Investigating the incidence of fluorinating enzymes in *Dichapetalum cymosum* and *Actinopolyspora mzabensis*

by

Selisha Ann Sooklal

(450050)

Thesis

Submitted in fulfilment of the requirements for the degree

Doctor of Philosophy

in

Molecular and Cell Biology

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Prof. Karl Rumbold

August, 2019

Declaration

I, **Selisha Ann Sooklal (450050)**, am a student registered for the degree of **Doctor of Philosophy** in the academic year 2019.

I hereby declare the following:

I. I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
II. I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
III. I have not submitted this work before for any other degree or examination at this or any other University.
IV. The information used in the Thesis HAS NOT been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.
V. I have followed the required conventions in referencing the thoughts and ideas of others.
VI. I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature: SA Sooklal Date: 05 August 2019

Dedicated to my parents, Peter and Sivy Sooklal,

and to my dear partner, Silviu Tomescu.

Strive for progress, not perfection.

-Anonymous

Investigating the incidence of fluorinating enzymes in *Dichapetalum cymosum* and *Actinopolyspora mzabensis*

Abstract

Despite fluorine being the most abundant halogen in the earth's crust, fluorinated metabolites are extremely rare in nature. The first fluorinated metabolite, fluoroacetate, was isolated from the plant *Dichapetalum cymosum*. Given the scarcity of fluorometabolites, the biosynthetic origin of fluoroacetate in *D. cymosum* was of great interest. However, the fluorination mechanism in plants was never elucidated and attention shifted to the bacterium *Streptomyces cattleya*, which was also found to produce fluoroacetate as well as 4-fluorothreonine. The first fluorinating enzyme, or fluorinase, was isolated from *S. cattleya*. Fluorinases have the unique ability to catalyse a C-F bond, hence, have vast potential as biocatalysts in the production of fluorinated products. But fluorinases are extremely rare enzymes and only five representatives have been isolated thus far. Therefore, efforts to expand enzymatic resources for biofluorination are highly desired. This study aimed to investigate *D. cymosum* on an omics level to gain insights into the potential fluorinating mechanism. *D. cymosum* presumably produces fluoroacetate as a defence mechanism, hence, mechanical wounding studies were performed and cDNA libraries of the unwounded (control) and wounded (30 and 60 minutes post wounding) plant were sequenced. The transcriptome assembly generated 77,845 transcripts, of which, 69% were annotated using various databases. This represents the first comprehensive transcriptome for *D. cymosum*. While the differential expression analysis revealed key wound-responsive transcripts related to signalling cascades, phytohormone regulation, transcription factors and some secondary metabolite defences; a transcript with homology to any known fluorinase was not detected. A functional screening based approach coupled with proteomic sequencing was then employed. A total of 2,430 proteins were detected. However, assays using the natural substrate of bacterial fluorinases did not yield any activity in *D. cymosum*. The lack of a transcript with homology to bacterial fluorinases and failure to detect activity using bacterial fluorinase substrates strongly suggests the presence of a different fluorinating enzyme in the plant kingdom. Efforts to expand fluorinase resources were

extended to bacteria. We report on the identification of a novel fluorinase from *Actinopolyspora mzabensis*. A hypothetical protein sequence, herein denoted as *Am*-fluorinase, was identified in the genome of *A. mzabensis*. Sequence alignments and homology models of this sequence revealed features characteristic of fluorinases, suggesting a promising new target. Thus, fluorination activity was validated experimentally. The gene was cloned into *E. coli* BL21 (DE3) cells. Overexpression of the *Am*-fluorinase resulted in the formation of insoluble inclusion bodies. A method for the recovery of the active enzyme from inclusion bodies is presented. Following purification, > 80 mg of the *Am*-fluorinase was obtained. Characterisation of the biochemical reaction confirmed fluorination activity. The optimal pH for activity was found at pH 7.2 while the optimal temperature for activity was found at 65 °C. At these conditions the enzyme exhibited a specific activity of 0.44 U/ mg. The presence of Mg^{2+} ions was also shown to increase relative activity by 10%. However, most intriguingly, the *Am*-fluorinase displayed exceptional stability at 25 °C indicating thermostability. The identification of a novel fluorinase in this study contributes to the biological tools available for fluorination. In addition, this is the first report on the successful refolding of a fluorinase enzyme and the method employed here can potentially be applied to other fluorinases to obtain high yields of the enzyme.

Keywords: *Dichapetalum cymosum*, transcriptome, proteome, *Actinopolyspora mzabensis*, fluorinase, C-F bond, biocatalyst

Research outputs

Publications forming part of PhD thesis:

1. Sooklal, S. A., Mpangase, P., Aaron, S., Archer, R. H. and Rumbold, K. (2019) Functional characterisation of the transcriptome from the fluoroacetate-producing plant, *Dichapetalum cymosum*, in response to mechanical wounding. *Scientific Reports*. Submitted for publication.

2. Sooklal, S. A. and Rumbold, K. (2019) Identification and characterisation of a novel fluorinase from *Actinopolyspora mzabensis*. *Protein Expression and Purification*. Submitted for publication.

Publications not forming part of PhD thesis:

1. Kassim, M. A., Sooklal, S. A., Archer, R. and Rumbold, K. (2014) Screening for hydroxynitrile lyase activity in non-commercialised plants. *South African Journal of Botany*. **93: 9-13**

Conference presentations:

1. Sooklal, S. A. and Rumbold. K. (2018) The identification and characterisation of a novel fluorinating enzyme from *Actinopolyspora mzabensis.* Wits 9th Cross-Faculty Postgraduate Symposium. Oral presentation.

2. Sooklal, S. A. and Rumbold, K. (2017) Investigating the fluorination pathway in the toxic South African plant, *Dichapetalum cymosum*. Catalysis Society of South Africa 2017. Oral presentation.

3. Sooklal, S. A., Mpangase, P., Aron, S. and Rumbold, K. (2015) *De novo* assembly and characterization of the transcriptome from the South African fluoroacetate producing plant, *Dichapetalum cymosum*. Catalysis Society of South Africa 2015. Oral presentation.

Acknowledgements

Prof. Karl Rumbold, thank you for your continuous guidance and motivation throughout the many years I have spent in the IMBL lab. You have always given me the opportunity to choose my own direction with this project and it has allowed me to develop as a scientist. I have thoroughly enjoyed working on the fluorinase project.

Dr. Robert H. Archer, thank you for always welcoming us at the Pretoria National Botanical Gardens whenever we needed to collect plant material.

Phelelani Mpangase and Shaun Aaron, I sincerely thank you for assisting in the bioinformatic analysis of the *D. cymosum* transcriptome.

Prof. Ruth Birner-Grünberger and Barbara Darnhofer, thank you for generously sequencing the *D. cymosum* proteome.

Dr. Allan Prior, thank you for kindly synthesising and providing the 5'-FDA sample.

Wits Mass Spec Facility, I thank the Wits Mass Spec Facility for access to state-of-the-art equipment. A sincere thank you, more especially, to Thapelo Mbele and Eric Morifi for training on the LC-MS.

National Research Foundation, thank you for funding (PhD grant no. 95062 and Extension support grant no. 111329).

My parents, Peter and Sivy Sooklal, all your sacrifices have allowed me to accomplish this achievement. Thank you for your endless encouragement and support. You always gave me comfort knowing that if I win or fail, you would always be there.

Silviu Tomescu, no words can express how grateful I am to have you in my life. You were the one person always there, in the lab and out. Thank you for supporting me throughout all the late nights and comforting me when science unkind.

Table of contents

List of figures

List of tables

List of abbreviations

4-FT	4-Fluorothreonine
5'-FDA	5'-fluoro-5'-deoxyadenosine
5'-FDRP	5-fluoro-5-deoxyribose-1-phosphate
5'-FDRu1P	5-fluoro-5-deoxyribulose-1-phosphate
5'-FHPA	5'-fluoro-2,3,4-trihydroxypentanoic acid
Å	Angstrom
bp	Base pair
CNS	Central nervous system
CoA	Coenzyme A
COG	Clusters of orthologous groups
cDNA	Complementary deoxyribonucleic acid
DEGs	Differentially expressed genes
eV	Electron volt
ET	Ethylene
EC	Enzyme commission
ESI	Electron spray ionisation
eggNOG	evolutionary genealogy of genes: Non-supervised Orthologous Groups
FA	Fluoroacetate
Ppm	Parts per million
SAM	*S*-adenosyl-$_L$-methionine
SAH	*S*-adenosylhomocysteine
TCA	Tricarboxylic acid
PLP	Pyridoxal 5'-phosphate
kDa	Kilo Daltons
NADH	Nicotinamide adenine dinucleotide
NMR	Nuclear magnetic resonance

ORFs	Open reading frames
PNP	Purine nucleoside phosphorylase
k_m	Michaelis-Menten constant
k_{cat}	Turnover number
tRNA	Transfer ribonucleic acid
PET	Positron emission tomography
RNA	Ribonucleic acid
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
LC-MS	Liquid chromatography-mass spectrometry
NGS	Next generation sequencing
FDR	False discovery rate
nr	non-redundant database
rRNA	Ribosomal ribonucleic acid
mRNA	Messenger ribonucleic acid
GO	Gene ontology
KEGG	Kyoto Encyclopedia of genes and genomes
SSR	Simple sequence repeat
PlantTFDB	Plant transcription factor database
PE	Paired-end
JA	Jasmonic acid
SA	Salicylic acid
ABA	Abscisic acid
IAA	Auxin
GA	Gibberellin
MWCO	Molecular weight cut-off
MRM	Multiple reaction monitoring
ROS	Reactive oxygen species

V	Volts
qPCR	Quantitative polymerase chain reaction
MS	Mass spectrometry
m/z	mass-to-charge ratio
IBs	Inclusion bodies
HIC	Hydrophobic interaction chromatography
SEC	Size exclusion chromatography

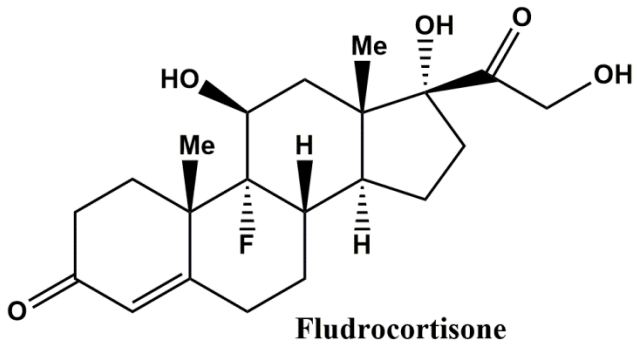
Chapter 1

Introduction

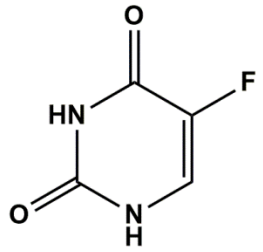
1.1 The significance of fluorinated products in industry

Fluorine has played a crucial role in history with the industrial production of fluorochemicals dating back to the late 1930s [1]. At the time, however, it was inconceivable to introduce fluorine into molecules with biological application since literature had suggested rather poisonous properties of fluorinated compounds. During a systematic study in 1953, Fried and Sabo had noticed that the bioactivity of cortisone had increased with specific halogenated derivatives [2]. This prompted an investigation into the fluorinated derivative which, surprisingly, showed more than 10-fold activity in comparison to the parent hormone [3]. This compound later became known as Fludrocortisone (Figure 1), a corticosteroid used to treat adrenal gland disorders, and represented the first fluorine-containing pharmaceutical to be introduced onto the market [4]. A few years later, 5-fluorouracil (Figure 1) was synthesized and presented as an antimetabolite of uracil [5]. Further studies had demonstrated the remarkable antitumor activity of 5-fluorouracil, a drug still widely used in cancer therapy today [6].

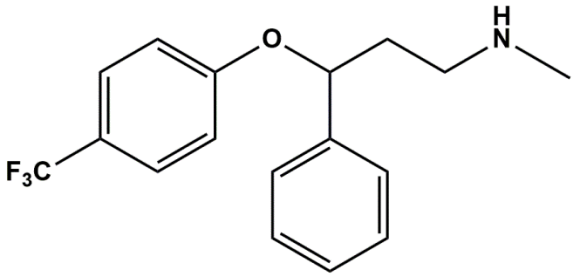
The discovery of Fludrocortisone and 5-fluorouracil led to a paradigm shift regarding the use of fluorine and, by 1970, approximately 2% of all pharmaceuticals contained a fluorine substituent [4,7]. This number has only increased with 25% of fluorine-containing drugs currently on the market; many of which are among the best-selling drugs such as atorvastatin (Lipitor) used for the treatment for high cholesterol levels, fluoxetine (Prozac) used as an antidepressant and ciprofloxacin (Ciprobay) used as an antibiotic (Figure 1) [4,8]. Indeed, the same trend has also been observed with agrochemicals and it has been estimated that more than 25% of herbicides contain fluorine [8,9].



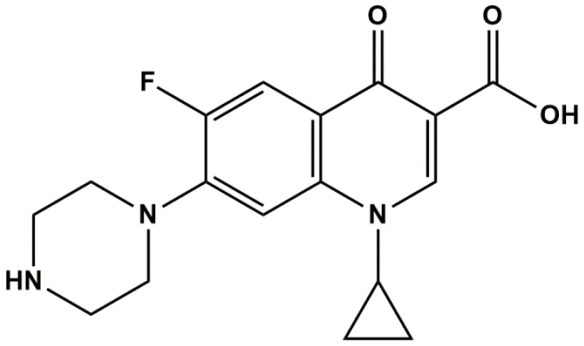
Fludrocortisone



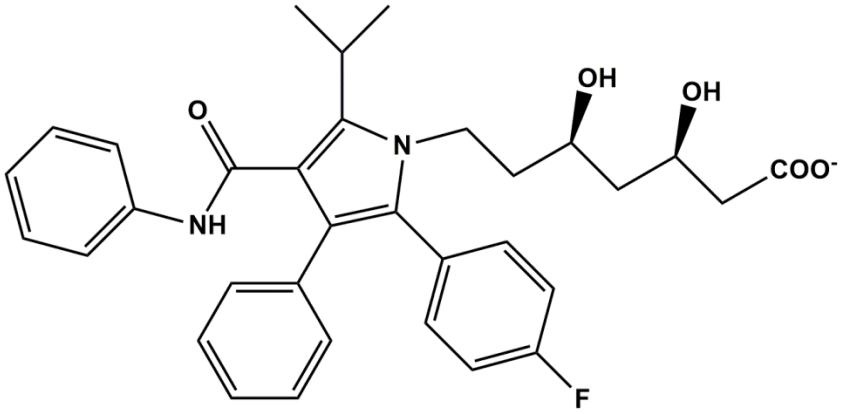
5-fluorouracil



Fluoxetine (Prozac)



Ciprofloxacin (Ciprobay)



Atorvastatin (Lipitor)

Figure 1: Chemical structures of fluorinated pharmaceuticals. Fludrocortisone and 5-fluorouracil were some of the first fluorinated pharmaceuticals introduced onto the market while fluoxetine, ciprofloxacin and atorvastatin are amongst the best-selling drugs currently on the market. Structures were produced using ChemDraw version 12.0.2.

The appeal in fluorination emanates from the desirable qualities that fluorine has been shown to impart on molecules. This has included attributes such as distinct electronic properties, enhanced drug potency and improved chemical or metabolic stability; allowing fluorine to find its way into various industries, most prominently in pharmaceuticals and agrochemicals [10]. These attributes are readily explained by the unique properties of fluorine.

Apart from having a small atomic radius that is only slightly larger than hydrogen, fluorine is also the most electronegative element and forms the strongest single bond with carbon [11,12]. This makes fluorine a suitable substituent in lieu of hydrogen. Moreover, the C-F bond length (1.39 Å) is in-between that of C-H (1.09 Å) and C-O (1.43 Å) providing minor steric effects. However, the electron withdrawing properties of fluorine results in a strong polarisation of the C-F bond [8,12]. Although there are minimal steric effects, this implicates major electronic effects. Polarisation of the C-F bond can alter the acidity of neighbouring functional groups and electrostatic interactions can lead to conformational changes in small molecules [13]. This directly influences binding affinities, pharmacokinetics and bioavailability of drug candidates [4,10]. The introduction of a C-F bond has been shown, in some cases, to modulate the lipophilicity of organic molecules due to the tendency of water molecules to order around the C-F unit. Based on the position of the C-F bond, this allows for the tuning of affinities to protein receptors [10,14]. Modulation of metabolic stability is also a well-established effect of fluorination. Owing to the inertness of the C-F bond to biological oxidation by Cytochrome P450 monooxygenases, key enzymes in drug metabolism, the dose and duration of pharmaceutical agents are impacted [10,15].

Although the effect of fluorine on molecules is unpredictable, researchers have adopted fluorine scans as a routine approach in drug design given the potentially beneficial characteristics that this element may provide [4]. Furthermore, taking into account the success of fluorine in various industries, it is likely that the number of commercialised fluorinated products will continue to grow.

1.2 The scarcity of fluorinated products in nature

An overwhelming number of compounds with commercial value have been based on or directly derived from natural products [16,17]. But despite the industrial significance, this has not been the case for fluorinated products of value. Whilst there are more than 5000 natural products originating from other halogens such as chlorine, bromine and iodine; only 6 naturally occurring fluorinated metabolites (fluoroacetate, (2*R*, 3*R*)-2-fluorocitrate, ω-fluorooleic acid, nucleocidin, 4-fluorothreonine and 5'-fluoro-2,3,4-trihydroxypentanoic acid) have been securely identified thus far [18,19]. The dearth of natural fluorinated products is frequently attributed to the bioavailability of fluorine, the difficulty in desolvating the fluoride ion and exclusion from enzymatic incorporation [19].

1.2.1 Bioavailability

Despite fluorine being the most abundant halogen in the Earth's crust, it is most often found in an insoluble state such as fluorite (CaF_2) or cryolite (Na_3AlF_6). This inherently reduces the bioavailability of the fluoride ion (F^- = 1.3 ppm in oceans), relative to the chloride ion (Cl^- = 20,000 ppm in oceans) and the bromide ion (Br^- = 70 ppm in oceans), where it could possibly participate in reactions (Table 1) [19,20]. However, even though the iodide ion is far less abundant in oceans (I^- = 0.02 ppm), it has still given rise to ~ 120 natural products suggesting that desolvation of the fluoride ion and exclusion from enzymatic incorporation play a fundamental role in the scarcity of fluorinated natural products [21].

Table 1: Physical and chemical properties of F, Cl, Br and I. Adapted from [11].

Halogen	Electronegativity (Pauling scale)	Heat of hydration (kcal mol^{-1})	Redox potential (eV)	Relative abundance (ppm)	
				Earth crust	Ocean
F	4.0	120	- 3.06	540	1.3
Cl	3.0	84	- 1.36	170	20,000
Br	2.8	78	- 1.07	3	70
I	2.5	68	- 0.54	0.49	0.02

1.2.2 Desolvation

In its desolvated state, the fluoride ion is a potent nucleophile. However, due to its electronegativity, the fluoride ion has a high propensity for hydration establishing strong hydrogen bonds with water molecules [11]. Indicated by the highest heat of hydration of all the halogens (~120 kcal.mol^{-1}), as shown in Table 1, there is a considerable energy cost that is required in order to desolvate the fluoride ion before it can participate in displacement reactions [19]. This has resulted in the fluoride ion being relatively inert in biological systems. The evolution of an enzyme with an appropriate desolvation strategy would be the only way to achieve nucleophilic catalysis *in vivo*.

1.2.3 Enzymatic incorporation

In biological systems, halogen-containing molecules are primarily a product of enzyme-catalysed halogenation [21]. A diverse group of enzymes, halogenases, catalyse the addition of halogens (Cl^-, Br^- and I^-) onto organic molecules [22]. The first halogenating enzyme was identified in the fungus *Caldariomyces fumago* during investigations into the biosynthesis of caldariomycin, a chlorinated metabolite. In the presence of H_2O_2, the enzyme was able to catalyse the addition of chloride ions onto an intermediate molecule in the caldariomycin biosynthesis pathway [23,24]. Hence, the enzyme was termed a chloroperoxidase. The chloroperoxidase from *C. fumago* was shown to utilise bromide and iodide ions as well [25]. Development of an assay for the characterisation of this enzyme allowed for the subsequent detection of chloroperoxidases (utilise Cl^-, Br^- and I^-), bromoperoxidases (utilise Br^- and I^-) and iodoperoxidases (utilise I^-) from a variety of prokaryotes and eukaryotes [25-30]. The aforementioned haloperoxidases catalyse the vast majority of halogen modifications seen in natural products. However, the inability of these enzymes to utilise fluoride ions became a trend due to limitations in the oxidative mechanisms employed by these enzymes [21,22]. These mechanisms are better explained by describing the three oxidation states exploited by halogenases to facilitate halide incorporation.

1.2.3.1 Halogenation by X^+

Halogenases that use this mechanism facilitate the two-electron oxidation of a halide ion (X^-) to form a halenium ion (X^+). The halenium ion is then involved in an electrophilic substitution reaction with the corresponding substrate [22]. A prerequisite, however, is that substrates are electron-rich allowing for a nucleophilic character in order to react with the halenium ion. This mechanism is carried out in the presence of H_2O_2 which provides the redox potential necessary to generate the oxidised equivalent of each halide [22]. This mechanism represents the most common enzymatic halogenation strategy and is observed in vanadium-dependent haloperoxidases [26,27,31] and heme-dependent haloperoxidases [32,33]. An alternative strategy is employed by flavin-dependent halogenases. These enzymes rely on O_2 instead of H_2O_2 to provide the redox potential, however, still adhere to the two-electron oxidation mechanism [34,35]. Halogenases in this group have only been found to catalyse reactions where $X^- = Cl^-$, Br^- or I^-.

1.2.3.2 Halogenation by $X^\bullet$

The previous mechanism for halogenation could not account for natural products that were halogenated on unactivated alkyl carbon atoms [22]. This led to the identification of an unconventional mechanism that employed radical chemistry. Halogenases in this group, or α-ketoglutarate-dependent non-heme iron halogenases, mediate the one-electron oxidation of a halide ion (X^-) to form a $X^\bullet$ radical, and halogenate unactivated aliphatic carbon centres [36]. This mechanism is carried out in the presence of O_2 and is dependent on α-ketoglutarate and Fe(II) for halogenation [37]. These halogenases have also only been found to catalyse reactions where $X^- = Cl^-$, Br^- or I^-.

Both halogenation by X^+ and halogenation by $X^\bullet$ rely on the oxidation of the halide ion either using H_2O_2 or O_2. Enzymes in these categories readily oxidise chloride (- 1.36 eV), bromide (-1.07 eV) and iodide (-0.54 eV) ions, but not the fluoride ion (- 3.06 eV) since the redox potential generated by hydrogen peroxide (- 1.8 eV) is insufficient (Table 1) [11,19]. Therefore, these mechanisms automatically exclude the fluoride ion [22]. As previously stated, even though the iodide ion is less abundant in oceans paralleled to the fluoride ion, it has still given rise to ~ 120 natural iodine-containing products. This is essentially due to the ease in which the iodide ion is oxidised [19, 21].

1.2.3.3 Halogenation by X^-

In contrast to the previous groups, halogenases in this class use the halide ion (X^-) directly as nucleophiles to catalyse S_N2 reactions on an electron-deficient carbon centre, eliminating the need for prior oxidation [22]. A unifying feature of nucleophilic halogenases is the utilisation of *S*-adenosyl-$_L$-methionine (SAM), an activated electrophile, as a substrate. Based on the diversity of halogenated products in nature, this mechanism for halogenation is rather infrequent and is occupied by only halide methyltransferases and 5'-halo-5'-deoxyadenosine synthases [38-40].

Halide methyltransferases are responsible for methyl halides observed in some fungi, algae and plants. These enzymes catalyse the nucleophilic substitution between halide ions (X^-) and SAM which results in the generation of a methyl halide and the displacement of *S*-adenosylhomocysteine (SAH) [38,41,42]. Due to its electronegativity, removal of an electron from the fluoride ion is highly unfavourable, discounting previous oxidation mechanisms of halogenases for F^- incorporation [22]. Halogenation by X^- would therefore seem as the only logical mechanism for F^- integration since the fluoride ion has the potential to be a potent nucleophile in solution, provided that it is desolvated first. However, all known halide methyltransferases have been shown to accept only Cl^-, Br^-, I^- and other nucleophiles such as thiocyanate and bisulfide [38,41-44]. This indicated that most halogenases also do not have the capacity to desolvate the fluoride ion prior to catalysis. But documentation of naturally occurring fluorinated metabolites, such as fluoroacetate, challenged the notion of complete exclusion of fluorine from enzymatic reactions. This suggested the evolution of some mechanism to overcome the difficulty in desolvating the fluoride ion.

The discovery of 5'-halo-5'-deoxyadenosine synthases consolidated a major gap in knowledge regarding the natural incorporation of fluorine in prokaryotes [40]. Although exceedingly rare enzymes, 5'-halo-5'-deoxyadenosine synthases catalyse the nucleophilic attack on the C-5' carbon of SAM to generate 5'-halo-5'-deoxyadenosine and $_L$-methionine [45]. This group is pioneered by the fluorinase from *Streptomyces cattleya* and the chlorinase from *Salinispora tropica* [39,40]. Fluorinases represent the only enzymes capable of catalysing a C-F bond; however, these enzymes will be discussed in more detail in section 1.4 Fluorinated metabolites in bacteria.

The low bioavailability of fluorine, high energetic cost for desolvation of the fluoride ion and exclusion of fluorine from enzymatic incorporation via most halogenases has severely restricted the natural evolution of fluorine biochemistry. Consequently, only six fluorinated natural products (fluoroacetate, (2*R*, 3*R*)-2-fluorocitrate, ω-fluorooleic acid, nucleocidin, 4-fluorothreonine and 5'-fluoro-2,3,4-trihydroxypentanoic acid) have been securely identified to date and the field of fluorination, as we understand it, is completely synthetic.

1.3 Fluorinated metabolites in plants

1.3.1 *Dichapetalum cymosum*

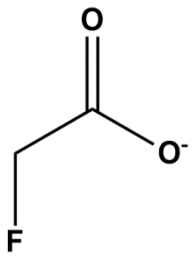
1.3.1.1 Distribution and morphology

Dichapetalum cymosum (Hook) Engl., or more commonly known as gifblaar (Afrikaans for poison leaf), is an extremely toxic plant found in South Africa, Namibia, Zimbabwe, Botswana and Angola [46]. Preferring deep sandy soils, *D. cymosum* is mainly scattered in northern regions of the Mpumalanga, Gauteng, Limpopo and North West provinces in South Africa [47]. This perennial shrub is characterised by an extensive underground stemming system with apical branches protruding the ground bearing bright green leaves. *D. cymosum* produces small white flowers which, on very rare occasions, develop into a round yellow-orange fruit (Figure 2) containing one to three seeds [48].

Figure 2: Photograph showing the bright green leaves and yellow-orange fruit of *D. cymosum*. Image obtained from Dr. Robert H. Archer.

1.3.1.2 Toxicity

After observing sudden death in animals that ingested *D. cymosum*, Marais extracted a compound termed 'potassium cymonate' from leaves of this shrub. Upon administration (orally, subcutaneously or intravenously), the compound exhibited high toxicity leading to death within a few hours in rabbits (5 $mg.kg^{-1}$ body weight) [49]. In 1944, Marais successfully identified monofluoroacetate, or fluoroacetate (FA), Figure 3, as the compound responsible for toxicity in *D. cymosum* [50]. However, more importantly, this discovery had documented the first fluorinated metabolite from a biological source.



Fluoroacetate

Figure 3: Chemical structure of fluoroacetate. Fluoroacetate was isolated as the toxic principle of *D. cymosum*. Structures were produced using ChemDraw version 12.0.2.

D. cymosum was subsequently shown to accumulate FA at concentrations up to 232 $mg.kg^{-1}$ fresh mass in young leaves. Flowers and immature seeds were also found to be some of the most toxic parts of the plant containing FA at 362 $mg.kg^{-1}$ and 410 $mg.kg^{-1}$ fresh mass. Lower concentrations of FA were found in the root (70 $mg.kg^{-1}$ fresh mass), stem (84 $mg.kg^{-1}$ fresh mass) and older leaves (97 $mg.kg^{-1}$ fresh mass) [51]. Apart from variations in FA concentration seen with different plant tissues, variations were also noted according to season with the uppermost toxicity observed in spring [52].

In the last survey of the economic impact of poisonous plants in South Africa, Kellerman reported that *D. cymosum* accounted for 8% of cattle mortality [47]. The lethal dose of FA has been estimated at 0.05 $mg.kg^{-1}$ for dogs; 0.4 - 2 $mg.kg^{-1}$ for cattle, goat, sheep and pigs; and 2 - 10 $mg.kg^{-1}$ for humans [53]. Toxicity of FA is effectuated through 'lethal synthesis' [54]. Once ingested, FA is metabolised *in vivo* to fluoroacetyl-CoA through the action of an acetyl-CoA synthase. Fluoroacetyl-CoA is then converted to (2*R*, 3*R*)-2-fluorocitrate via a

citrate synthase (Figure 4). (2*R*, 3*R*)-2-fluorocitrate, coincidently, is the only stereoisomer of fluorocitrate that is a competitive inhibitor of aconitate hydratase, a key enzyme in the tricarboxylic acid (TCA) cycle. This results in termination of the TCA cycle, a mechanism central to cellular respiration [55,58]. Furthermore, fluorocitrate has also been implicated in covalently binding citrate carrier proteins inhibiting mitochondrial transmembrane citrate transport and, thus, leading to citrate accumulation and acidosis [59,60]. Animals that ingest FA exhibit diverse symptoms. In general, dogs have displayed central nervous system (CNS) distress and respiratory failure which leads to death, while other grazing animals (rabbit, sheep, goat, cattle and horse) predominantly display ventricular fibrillation and cardiac failure. Pigs and humans, in contrast, have shown both respiratory failure and cardiac arrest [53,61-64].

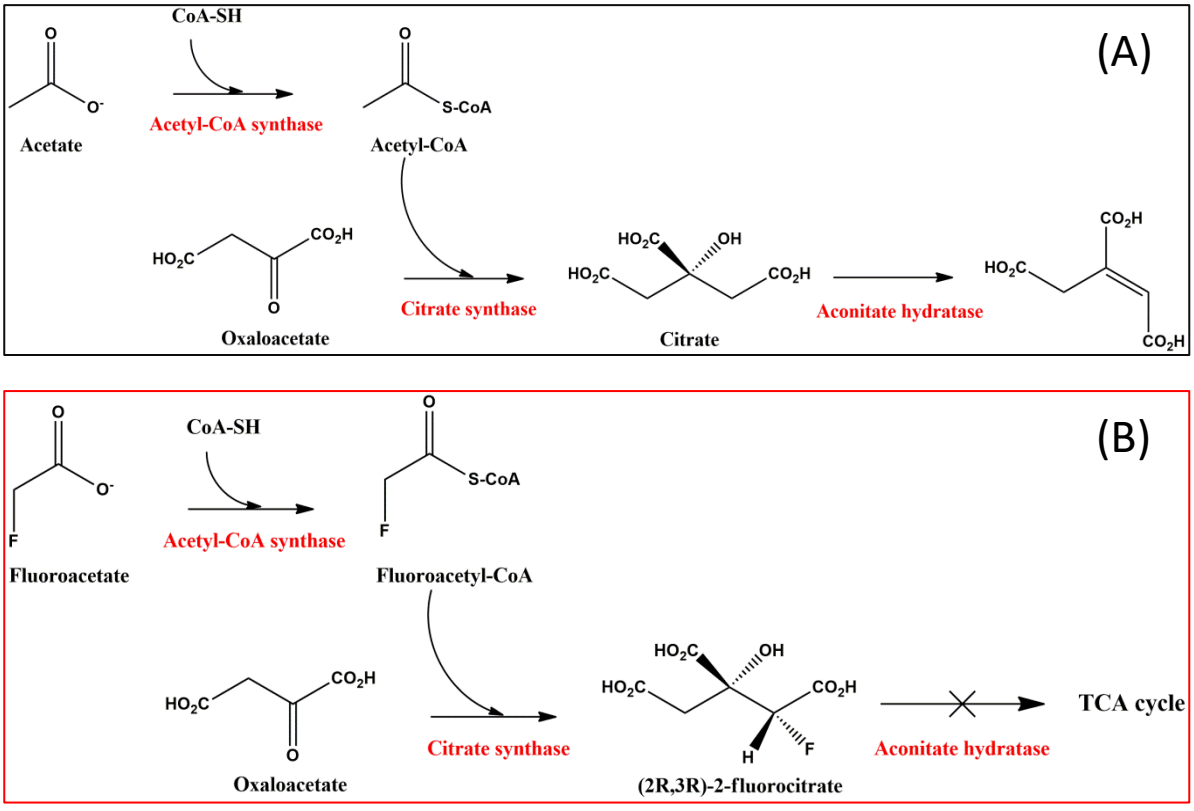


Figure 4: The mode of toxicity for fluoroacetate. Under normal conditions, the TCA cycles proceeds as shown in (A). However, when ingested, fluoroacetate is metabolised into (2*R*, 3*R*)-2-fluorocitrate as shown in (B). (2*R*, 3*R*)-2-fluorocitrate is an inhibitor of aconitate hydratase resulting in termination of the TCA cycle. Structures produced using ChemDraw version 12.0.2.

1.3.2 Other fluorometabolite-producing plants

Following the isolation of FA from *D. cymosum*, fluorometabolites were detected in other plants dispersed in tropical and semi-tropical regions of Africa, Australia and South America.

1.3.2.1 Africa

The incidence of fluorometabolite-producing plants in Africa appears to be confined to the family *Dichapetalaceae*. While *D. cymosum* is the only representative of the genus *Dichapetalum* in South Africa, several other FA-producing *Dichapetalum* species populate the rest of Africa including *D. braunii*, *D. toxicarium*, *D. heudelotti*, *D. struhlmannii*, *D. michelsonii*, *D. barteri* and *D. edule* [50,62,65-69]. There is also strong evidence for the presence of FA in *D. macrocarpum, D. mossambicense*, *D. schleibenii*, *D. barbosae*, *D. guineense* and *D. venenatum* [70]. Interestingly, *D. braunii* contains the highest levels of FA (7200 $mg.kg^{-1}$ dry mass in young leaves and 8000 $mg.kg^{-1}$ dry mass in seeds) seen in plants thus far [65]. Moreover, additional fluorometabolites, ω-fluorooleic acid and a series of derivatives (Figure 5), were shown to accumulate in the seed of *D. toxicarium* [71,72] These fluorinated lipids presumably originate from FA, where fluoroacetyl-CoA is thought to be the starter unit in fatty acid biosynthesis [19]. However, fluorinated lipids are exclusive to *D. toxicarium* and have not been detected in either *D. cymosum* or *D. braunii*. *Tapura fischeri*, the only representative outside the genus *Dichapetalum* in Africa, was also confirmed to produce FA [73].

1.3.2.2 Australia

It has been estimated that livestock poisoning and death, due to fluoroacetate-producing plants, costs the Australian livestock industry ~ 45 million AUD dollars annually [74]. FA has been detected in a single species from the genus *Acacia* (*A. georginae*) and ~ 31 species from the genus *Gastrolobium* [75-80] Many FA-producing species, previously described from the genus *Oxylobium*, have been moved to the genus *Gastrolobium* due to recent molecular phylogenetic studies [80].

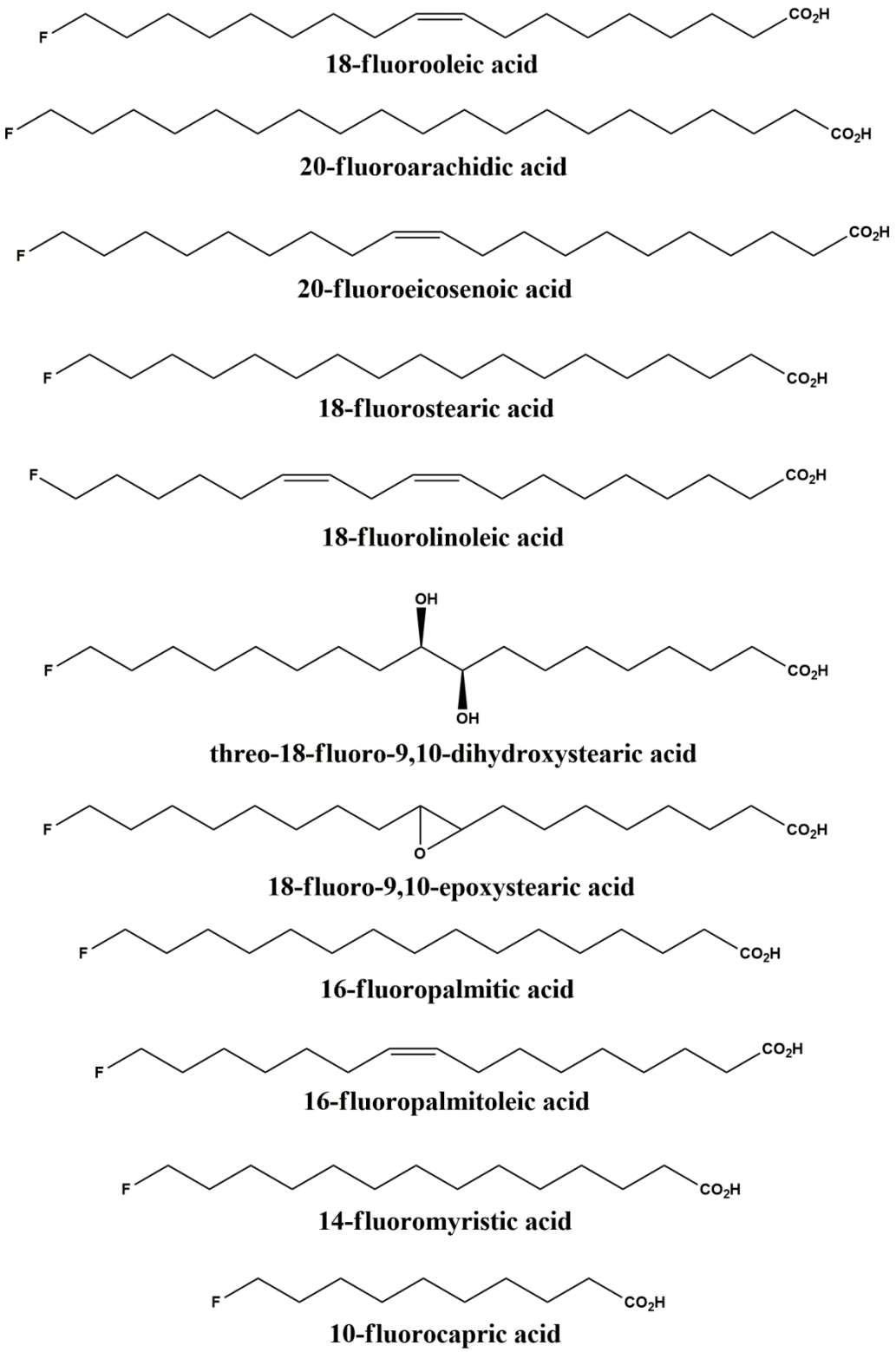


Figure 5: Chemical structures of fluorinated fatty acids isolated from *D. toxicarium*, a fluorometabolite-producing plant. Structures produced using ChemDraw version 12.0.2.

1.3.2.3 South America

FA-producing plants are responsible for ~ 60% of the total cattle loss attributed to poisonous plants in Brazil [81]. While *Amorimia* and *Palicourea* are the major FA-producing genera in South America with 16 species collectively, a single species from *Tanaecium* (*T. bilabiatum* or previously known as *Arrabidaea bilabiata*) was also confirmed to contain fluoroacetate [82-84].

1.3.3 The mystery of FA production in plants

The biogenesis and accumulation of FA is widely believed to be a defence mechanism in plants [74]. However, given how rare fluorometabolites are in nature, there was far more interest in the biosynthetic origin of these molecules. This led to vast speculation and many attempts to elucidate the fluorometabolite biosynthetic mechanism in higher plants.

In 1972, Hall and Cain proposed that *D. cymosum* acquires FA from microorganisms in its rhizosphere after detecting trace amounts of FA in the soil [85]. Although the acquisition of FA by *D. cymosum* was successfully demonstrated using ^{14}C-fluoroacetate, Meyer and Grobbelaar argued that the presence of FA in the surrounding soils could be due to the release of the metabolite from decomposing plant material [51]. Meyer and Grobbelaar found that, instead, microorganisms colonise *D. cymosum* internally which led to the conjecture of a FA-producing endosymbiont [86]. Grobbelaar and Meyer subsequently manufactured an aseptic growth chamber for *D. cymosum* seedlings. The seedlings were able to biosynthesise FA, however, the authors were unable to determine if the initial seeds were internally aseptic. Therefore, a callus culture of *D. cymosum* which appeared to be devoid of any internally contaminating organisms was established. The *D. cymosum* callus retained the ability to produce FA for many years when supplemented with fluoride ions [87]. Moreover, similar results were seen in *A. georginae*. A tissue culture of the stem was able to synthesise FA in the presence of fluoride ions [88]. These findings suggested the evolution of an enzyme with the capacity to desolvate and incorporate fluoride ions into organic molecules.

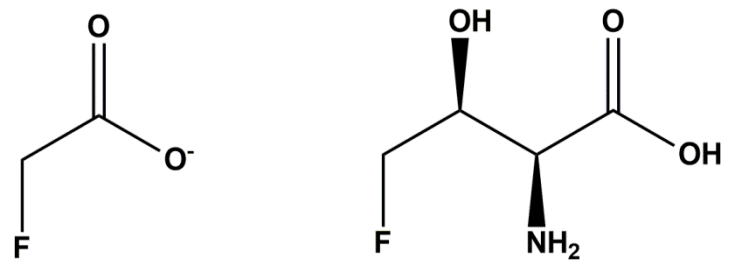
Further attempts to clarify the mechanism for fluorination in *D. cymosum* were challenged by access to plant material as well as by a general lack of knowledge regarding the natural substrate of a potential fluorinating enzyme [87,89,90]. Therefore, despite the effort, very

little progress was achieved. Attention transiently shifted to a fluorometabolite-producing bacterium, *Streptomyces cattleya*, providing a more practical system to study the fluorination pathway. The mechanism has not been elucidated in higher plants and remains a mystery to date.

1.4 Fluorinated metabolites in bacteria

1.4.1 *Streptomyces cattleya*

During fermentation studies to improve the production yields of the antibiotic thienamycin, Sanada and co-workers found that *S. cattleya* was also able to synthesize FA and 4-fluorothreonine (4-FT) when supplemented with fluoride ions (Figure 6) [91]. Of the three *S. cattleya* strains tested (NRRL 8057, MA 5176 and MA 5617), all were confirmed to produce FA using ^{19}F-NMR. However, only two (NRRL 8057 and MA 5176) of these strains had the ability to synthesize 4-FT. Synthesis of 4-FT and FA was observed during the stationary growth phase and supplementation with other halide ions (Cl^-, Br^- and I^-) did not yield the corresponding analogues of 4-FT [91]. These findings were momentous because this bacterium provided a more practical biological system, as compared to plants, to study the biosynthetic origin of fluorometabolites.



Fluoroacetate 4-fluorothreonine

Figure 6: Chemical structures of the fluorinated metabolites (fluoroacetate and 4-fluorothreonine) isolated from the bacterium *S. cattleya*. Structures produced using ChemDraw version 12.0.2.

1.4.1.1 Fluorometabolite biosynthesis pathway

After Sanada et al proposed that FA could be converted to 4-FT [91], subsequent studies focussed on establishing the relationship between these two metabolites. Using a variety of ^{2}H and ^{13}C isotopically labelled precursors, Reid et al showed that 4-FT was not synthesized from FA [92]. Further studies demonstrated that the integration of ^{13}C in the C-1 and C-2 of FA were identical to the C-3 and C-4 of 4-FT [93]. This suggested that both the fluorometabolites arose from a common C_2 precursor unit. Fluoroacetaldehyde was later found as the mutual precursor which accounted for labelling patterns observed in both FA and 4-FT [94].

Determination of fluoroacetaldehyde as the precursor, allowed for enzyme activities to be tracked in the biosynthesis of FA and 4-FT. A fluoroacetaldehyde dehydrogenase and a pyridoxal 5'-phosphate (PLP)-dependent transaldolase were thereafter purified from *S. cattleya* [95,96]. The fluoroacetaldehyde dehydrogenase mediates the conversion of fluoroacetaldehyde to FA in a NADH-dependent manner [95]. The PLP-dependent transaldolase was found to condense fluoroacetaldehyde and $_{L}$-threonine to form 4-FT [96].

Although the previous studies provided crucial information regarding the immediate precursor of FA and 4FT, it simply illustrated the conversion of a fluorinated intermediate and did not elucidate the initial substrate for fluorination. Hence, utilising high field ^{19}F-NMR, cell-free extracts of *S. cattleya* were screened using inorganic fluoride and an array of potential substrates [97]. These experiments revealed that *S*-adenosyl-$_{L}$-methionine (SAM) repeatedly elevated fluorometabolite production. It appeared that SAM may be converted to 5'-fluoro-5'-deoxyadenosine (5'-FDA). To verify these observations, a synthetic sample of 5'-FDA was used for comparison with the crude cell-free extract reaction product. This confirmed SAM as the initial substrate and 5'-FDA as the fluorinated reaction product [98]. Knowledge of the substrate and reaction product permitted tracking of enzymatic activities. The enzyme 5'-fluoro-5'-deoxyadenosine synthase, or trivially termed the fluorinase, was subsequently purified from *S. cattleya* [99]. The enzyme was shown to catalyse the first committed step in fluorometabolite biosynthesis by converting inorganic fluoride ions and SAM to 5'-FDA and $_{L}$-methionine [40,99]. These findings were a significant breakthrough in understanding biological fluorination since the fluorinase from *S. cattleya* represented the first enzyme with the capacity to catalyse a C-F bond, overcoming limitations associated with the incorporation of the fluoride ions.

Later research aimed at linking the intermediates between 5'-FDA and fluoroacetaldehyde. The phosphorolysis product of 5'-FDA, 5-fluoro-5-deoxyribose-1-phosphate (5-FDRP), was established as the next intermediate in the pathway [100]. This led to the partial purification of a purine nucleoside phosphorylase (PNP) from *S. cattleya* with the capacity to convert 5'-FDA to 5-FDRP [100]. The identification of 5-fluoro-5-deoxyribulose-1-phosphate (5-FDRu1P) directed the isolation of an isomerase from *S. cattleya* that catalyses 5-FDRP to form 5-FDRu1P [101,102]. An aldolase that catalyses the cleavage of 5-FDRu1P to form fluoroacetaldehyde was proposed as the remaining step in the fluorometabolite biosynthesis pathway [102]. However, an aldolase purified from *S. cattleya* was shown to produce 5-fluoro-5-deoxyfructose-1-phosphate, instead of 5-FDRu1P, when monitoring the reverse reaction between fluoroacetaldehyde and dihydroxyacetone phosphate [103]. This suggested the enzyme employed a different stereochemical course to that needed for fluorometabolite biosynthesis and therefore was not the aldolase involved in the pathway. The aldolase for fluorometabolite production has still not been identified in *S. cattleya*. The fluorometabolite biosynthesis pathway is shown in Figure 7.

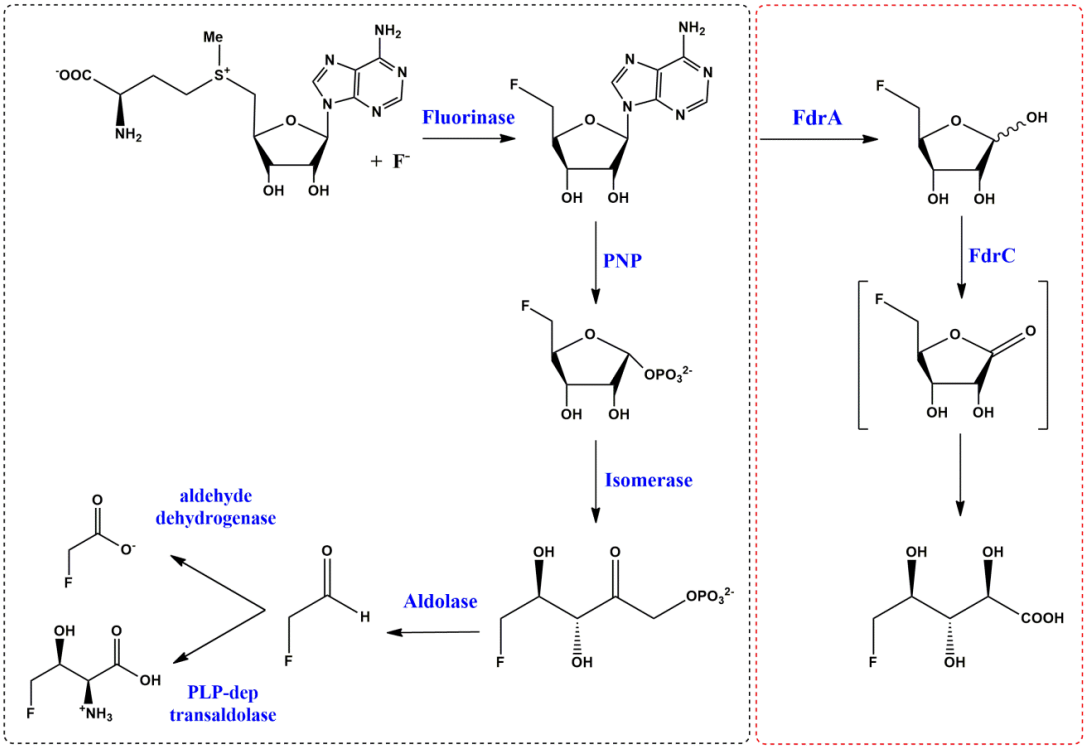


Figure 7: The fluorometabolite biosynthesis pathway. The pathway for the production of fluoroacetate and 4-fluorothreonine in *S. cattleya* is shown in the black box. A branch in this pathway (red box) was found in *Streptomyces* sp. MA37 which leads to the production of 5'-fluoro-2,3,4-trihydroxypentanoic acid.

1.4.1.2 Structure and mechanism of the fluorinase

The fluorinase, first purified from *S. cattleya*, represents the only class of enzyme able to catalyse a C-F bond. The enzyme is encoded by a 897 bp *fl*A gene [45]. This produces a polypeptide of 299 amino acids which corresponds to a 32 kDa monomeric subunit [45,99]. In solution, a native mass of approximately 180 kDa was observed suggesting the fluorinase forms a hexamer [99]. Subsequent cloning and recombinant expression of the fluorinase in *E. coli* enabled further structural studies.

The X-ray crystal structure of the fluorinase showed that each monomer contains two domains. The N-terminal domain, spanning from residues 8 to 180, consists of a central seven-stranded β-sheet sandwiched between α-helices. The C-terminal domain, spanning from residues 195 to 298, consists of a five-stranded and four-stranded antiparallel β-sheet [45]. The 22-amino acid (Lys92-Ala113) loop structure indicated in Figure 8A is a characteristic feature seen only in fluorinases. Three monomeric units arrange in a unique manner to form a trimer unit. The N-terminal domains form around a 3-fold axis and make contact with each other. The C-terminal domains, however, make contact with adjacent N-terminals but not with each other. SAM molecules bind at the interface between the N-terminal of one monomer and C-terminal of a neighbouring monomer (Figure 8). Each trimer can therefore bind three SAM molecules. The hexamer structure is formed by a dimer of trimers [45]. Structures of the monomer, trimer and hexamer are shown in Figure 8.

The desolvation of the fluoride ion has been one of the foremost limitations that have prevented incorporation of the element [11]. However, the fluorinase has evolved a strategy to overcome this limitation. The solvated fluoride ion discretely binds in a pocket defined from Phe156 to Ser158 making hydrogen bonds with the backbone NH and side chain OH of Ser158 [45]. In this process water molecules are exchanged, partially desolvating the fluoride ion and therefore minimising the energy cost for desolvation [45,104]. SAM is then coordinated into the active site binding with a higher affinity than the fluoride ion [104]. The binding of SAM further drives the desolvation of F^- as it is pushed into the hydrophobic pocket and the remaining water molecules dissociate. With F^- trapped and desolvated, it regains its potent nucleophilic character needed to attack the C-5' of SAM consistent with an S_N2 substitution reaction [45,104].

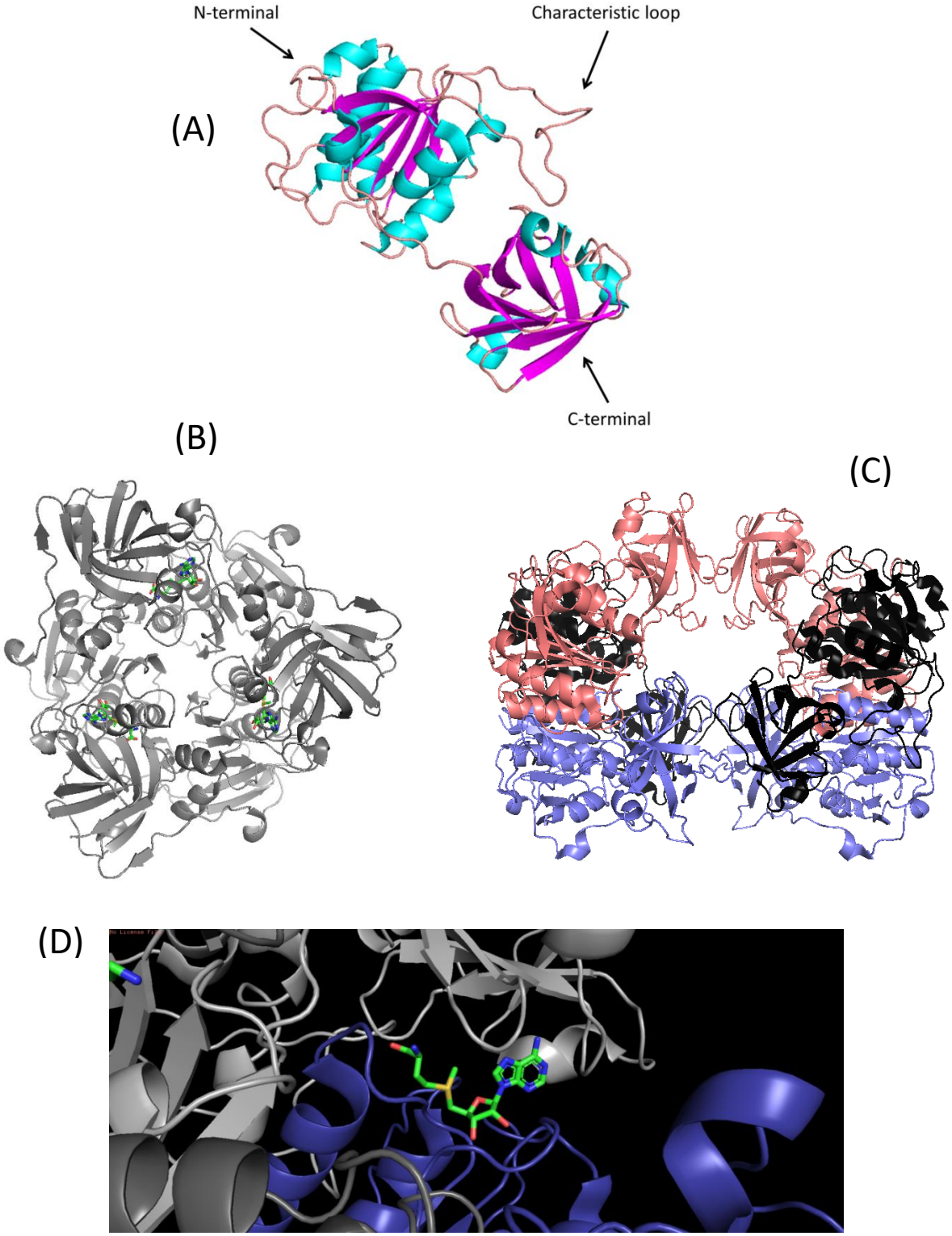
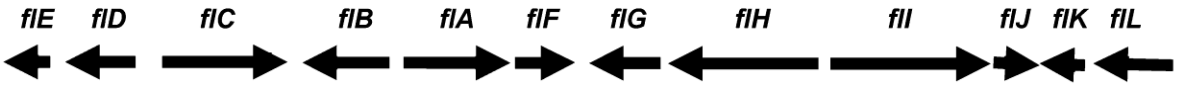


Figure 8: Structure of the fluorinase from *S. cattleya* (PDB 1RQP). (A) The monomer unit of the fluorinase consists of an N-terminal domain (residues 8-180) and C-terminal domain (residues 195-298). The loop structure characteristic of fluorinases is indicated by an arrow. (B) Trimer assembly of the fluorinase showing 3 SAM molecules (green) bound. (C) The native structure of the fluorinase is a hexamer formed by a dimer of trimers. (D) SAM molecule bound at the interface of two monomer subunits. Structures visualised using PyMol version 2.0.7.

1.4.1.3 Fluorometabolite gene cluster

Advances in sequencing technologies allowed for the full genome sequencing of *S. cattleya* as well as identification of the fluorometabolite gene cluster [103,105,106]. The fluorometabolite gene cluster, or *fl* cluster, is 12 kb in length and comprises of 12 open reading frames (ORFs) [106]. The organisation of genes in the *fl* cluster is shown in Figure 9.



Gene	Length (aa)	Putative Function
*fl*A	299	5'-FDA synthase (fluorinase)
flB	299	5'-FDA phosphorylase (PNP)
flC	397	Permease
flD	216	Dehalogenase
*fl*E	222	DNA binding regulatory protein
flF	185	DNA binding regulatory protein
flG	234	DNA binding regulatory protein
flH	467	Na^+/H^+ antiporter
flI	489	Adenosylhomocysteine hydrolase
flJ	131	Regulatory protein
flK	139	Fluoroacetyl-CoA thioesterase
*fl*L	225	DNA binding regulatory protein

Figure 9: Organisation and putative function of each ORF in the fluorometabolite (*fl*) gene cluster of *S. cattleya*. Image adapted from [106].

Interestingly, only 2 catalytic enzymes from the fluorometabolite biosynthesis pathway are encoded for in the *fl* cluster [103,106]. This includes the fluorinase (*fl*A) and PNP (*fl*B) which catalyses the first and second step in the fluorometabolite biosynthesis pathway, respectively. Disruption of the *fl*A gene in *S cattleya* resulted in the loss of ability to produce fluorometabolites. Similar results were seen with the PNP. Disruption of the *fl*B

gene also resulted in the loss of ability to produce fluorometabolites. This suggested that the fluorinase and PNP are the only enzymes that can perform their respective functions *in vivo* and play paramount roles in the pathway [103]. The remaining genes involved in the fluorometabolite biosynthesis pathway are scattered in remote locations of the *S. cattleya* genome [19].

Based on putative function, both *fl*C (permease) and *fl*H (Na^+/H^+ antiporter) are involved in transmembrane transport. The permease is thought to be involved in the transport of small molecules, such as fluorometabolites, across the cell membrane. However, this has not been confirmed as yet [106]. The Na^+/H^+ antiporter was initially proposed to have a role in facilitating F^- uptake but disruption of the *fl*H gene did not result in any loss of function. The *S. cattleya* mutant retained its full capacity to generate FA and 4-FT. This indicated that it is not essential for fluorometabolite biosynthesis and no significant role of the gene was realised [103].

The enzymes encoded for by the *fl*D (dehalogenase) and *fl*K (fluoroacetyl-CoA thioesterase) genes have putative roles in resistance. As previously mentioned, metabolic products of FA are toxic to living cells. The dehalogenase showed high homology to haloacid-dehalogenases and thus was proposed to function as a resistance mechanism, through defluorination of FA in *S. cattleya*. This has not been verified as yet. The fluoroacetyl-CoA thioesterase, on the other hand, was shown to convert fluoroacetyl-CoA to fluoroacetate and Co-A [106]. This reaction proceeded with remarkable selectivity for fluoroacetyl-CoA and discretely avoids the 'lethal synthesis' of (2*R*, 3*R*)-2-fluorocitrate [107]. Disruption of the *fl*K gene resulted in the loss of ability to produce fluorometabolites. This suggested regulatory aspects associated with the *fl*K gene which shuts down the fluorometabolite biosynthesis pathway when a loss of function is incurred [103].

The remaining genes in the *fl* cluster are proposed to have regulatory functions. The *fl*E, *fl*F, *fl*G and *fl*L genes are supposedly involved in transcriptional regulation while the regulatory function of the *fl*J gene is unknown. The *fl*I gene encodes an *S*-adenosylhomocysteine (SAH) hydrolase that catalyses the reversible hydrolysis of SAH [106]. SAH has been found as a competitive inhibitor of the fluorinase indicating a potential regulator of fluorinase activity *in vivo* [99,106].

1.4.2 Other fluorometabolite-producing bacteria

1.4.2.1 *Streptomyces* sp. MA37, *Actinoplanes* sp. N902-109, *Nocardia brasiliensis* and *Streptomyces xinghaiensis*

The fluorinase from *S. cattleya* persisted as the only representative for more than 10 years. However, more recently, 4 new enzymes were isolated.

The genomes of *Streptomyces* sp. MA37, *Actinoplanes* sp. N902-109 and *Nocardia brasiliensis* were deposited onto public databases. Through homology searches, putative fluorinases with > 80% similarity were identified (Figure 10). All of the sequences contained the 22 amino acid motif characteristic of fluorinases (Lys92-Ala113 *S. cattleya* numbering). Heterologous expression of these genes in *E. coli* confirmed novel fluorinases and characterisation of the enzymes revealed similar kinetic profiles to the fluorinase from *S. cattleya* (Table 3) [108,109].

Apart from FA and 4-FT, *Streptomyces* sp. MA37 exhibited the ability to produce another fluorometabolite. The fluorometabolite was identified as 5'-fluoro-2,3,4-trihydroxypentanoic acid (5'-FHPA), shown in Figure 11 [110]. This led to a branch in the fluorometabolite biosynthesis pathway (Figure 7). The production of 5'-FHPA seems to be an exclusive ability of *Streptomyces* sp. MA37 [110]. In contrast, *N. brasiliensis* failed to produce either FA or 4-FT in culture despite good cell growth. This suggested a latent pathway in *N. brasiliensis*. The fluorometabolite producing ability of *Actinoplanes* sp. N902-109 has not been determined to date since this bacterial strain is not publically available [108,109].

Similarly, the genome of *Streptomyces xinghaiensis* was deposited onto a public database [111]. A putative fluorinase with 84% identity to the *S. cattleya* fluorinase was identified (Figure 10). Heterologous expression in *E. coli* and characterisation of the enzyme verified a new fluorinase that was more efficient than previous fluorinases (Table 3) [112]. In culture, *S. xinghaiensis* was only able to produce FA. The presence of 4-FT was not detected [113]. Analysis of the genome revealed a truncated (PLP)-dependent transaldolase [19].

Given the recent advances in sequencing technologies and upsurge in the number of genomes available, fluorinases still remain relatively few [19]. This depicts the extremely limited distribution of these enzymes.

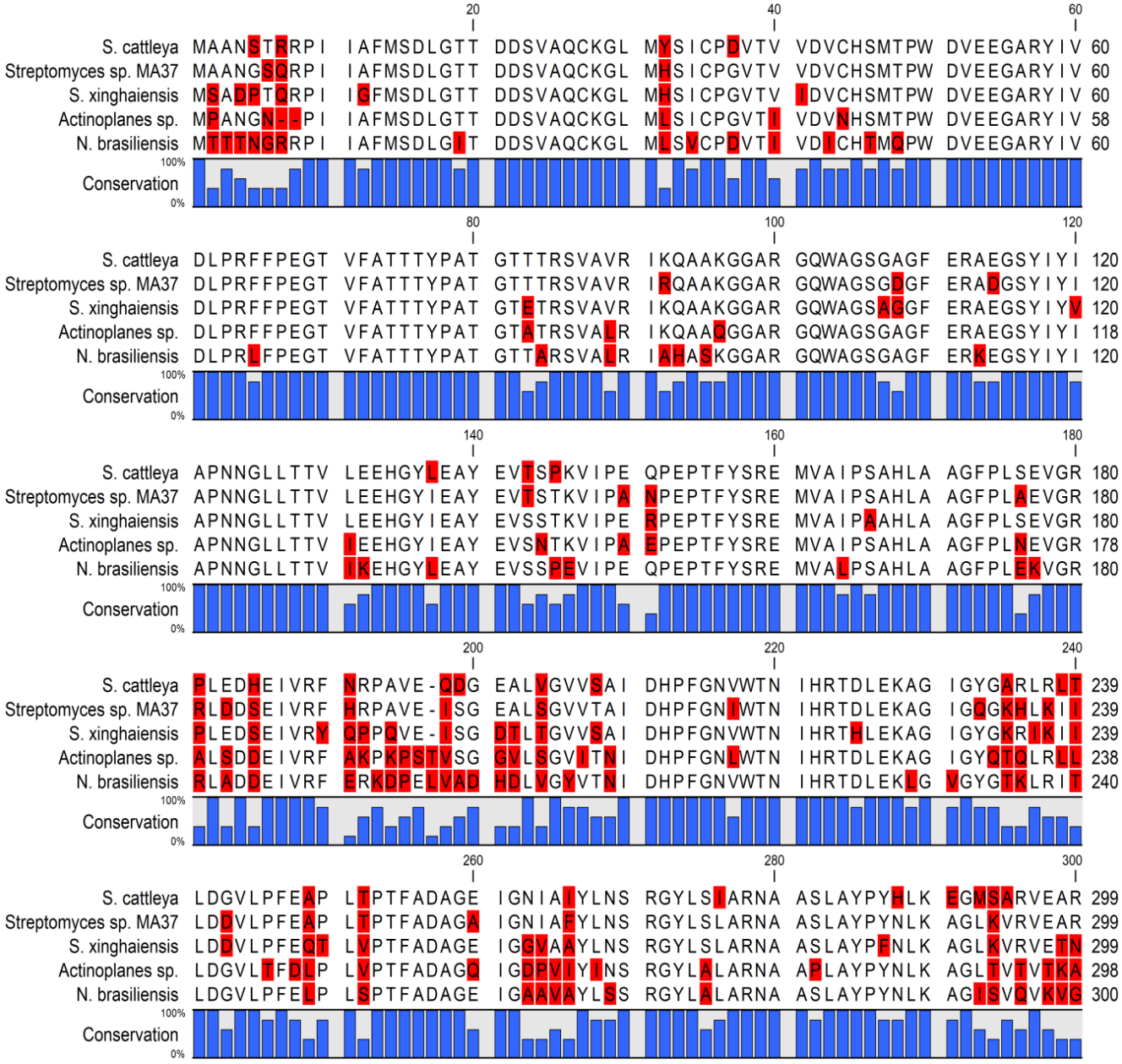
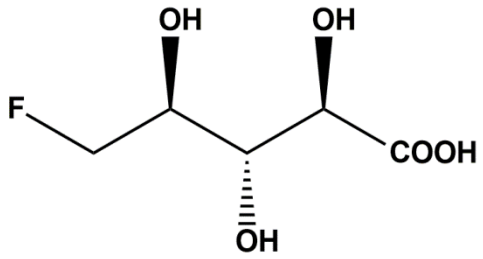


Figure 10: Multiple sequence alignment of the five known fluorinases. The fluorinase sequences for *Streptomyces cattleya* (WP_014144878.1), *Streptomyces* sp. MA37 (CDH39444.1), *Streptomyces xinghaiensis* (WP_019711456.1), *Actinoplanes* sp. N902-109 (WP_041832040.1) and *Nocardia brasiliensis* (WP_042263697.1) were retrieved from the GenBank database. The sequence alignment was performed on the CLC Genomics Workbench version 12.0 (Qiagen).

Table 2: Comparison of new bacterial fluorinases to the fluorinase from *S. cattleya*

Source	Identity (%)	SAM K_m (μM)	k_{cat} (min^{-1})	Reference
S. cattleya	-	29.2 ± 2.4	0.083	[99]
Streptomyces sp.	87	82.4 ± 18.6	0.262	[108]
Actinoplanes sp.	80	45.8 ± 7.91	0.204	[108]
N. brasiliensis	81	27.8 ± 4.23	0.122	[108,109]
S. xinghaiensis	84	7.04 ± 0.94	0.277	[112]



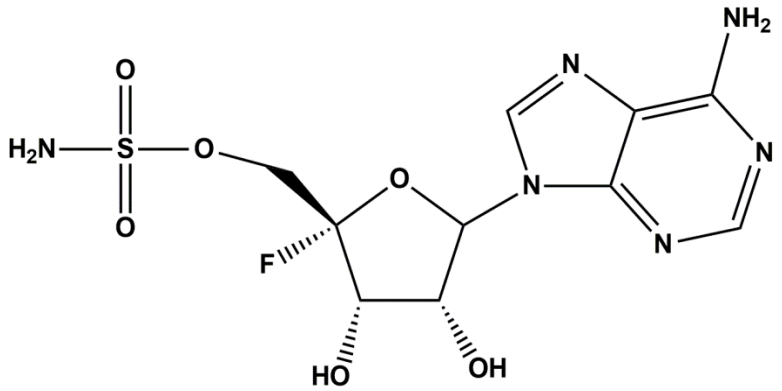
5-fluoro-2,3,4-trihydroxypentanoic acid

Figure 11: Chemical structure of an additional fluorinated metabolite (5'-fluoro-2,3,4-trihydroxypentanoic acid) isolated from *Streptomyces* sp. MA37. Structure produced using ChemDraw version 12.0.2.

1.4.2.2 *Streptomyces calvus*

Nucleocidin, a broad-spectrum antibiotic, was purified from the fermentation broth of *Streptomyces calvus* [114]. The correct structure of nucleocidin was later found to contain a fluorine atom and was identified as 4'-fluoro-5'-*O*-sulfamoyladenosine (Figure 12) [115]. Elucidation of the structure generated particular interest in this metabolite given the unique position of fluorine on the C-4 of the ribose moiety. This indicated that, unlike other fluorinated metabolites, nucleocidin was not derived from the same precursors. Efforts to reveal the mechanism were hindered by inability of *S. calvus* to reproduce nucleocidin under laboratory conditions [11]. Through genome sequencing of *S. calvus*, a point mutation was found in the *bldA* gene which encodes for a Leu-tRNA molecule [116].

Restoration of this gene has allowed *S. calvus* to regain its ability to produce nucleocidin [117]. Labelling studies revealed that nucleocidin may take a different course of biosynthesis to previous fluorometabolites [118]. A new pathway in this organism has not been found as of yet, however, a potentially new fluorinating mechanism is an exciting prospect.



Nucleocidin

Figure 12: Chemical structure of the fluorinated antibiotic (nucleocidin) isolated from *Streptomyces calvus*. Structure produced using ChemDraw version 12.0.2.

1.5 Biotechnological prospects of fluorinating enzymes

The traditional approach for fluorination has been through chemical synthesis. The vast majority of fluorinated products currently known are man-made. However, the incorporation of fluorine into organic molecules is often challenging under laboratory conditions [119]. This is primarily due to the electronegativity of fluorine. The high electronegativity results in a tight hydration shell around the fluoride ion mitigating its reactivity in aqueous solution. Removal of this hydration shell is necessary prior to catalysis, but requires a great deal of energy. Strong metal-fluoride bonds and highly polarised bonds further complicate incorporation of this element [119]. Hence, incorporation of fluorine often requires harsh conditions but still lacks selectivity. As such, alternative strategies that are safe and sustainable are highly sought after.

Biocatalysis, in contrast, is the utilisation of whole cells or enzymes in the synthesis of products. Enzyme-based catalysis offers regio-, chemo- and stereoselective production of

molecules. Moreover, enzymes are able to operate under mild conditions and are biodegradable. Biocatalysts, therefore, meet the demand for selective, safe and sustainable processes [120]. With the unique ability to catalyse a C-F bond, by utilising the fluoride ion as a nucleophile under mild conditions, fluorinases have vast potential as biocatalysts in biotechnological applications [19]. It is no surprise then that these enzymes have found their way into various applications.

Fluorinases have been applied in the synthesis of [^{18}F]-labelled ligands for positron emission tomography (PET). PET is a leading technology used for medical as well as biological imaging. Although [^{18}F]-fluoroacetate and a range of [^{18}F]-fluoronucleosides have been synthesised for PET imaging, the most promising compound so far was [^{18}F]-fluororibose [121-124]. This compound showed similar characteristics to [^{18}F]-2-fluoro-2-deoxyglucose, the imaging ligand most widely used [124]. Fluorinases have also been applied in synthetic biology approaches to produce fluorinated metabolites. Eustáquio et al demonstrated that the fluorinase gene can replace the chlorinase gene in *S. tropica* to produce fluorosalinosporamide, an analogue of the potent anti-cancer salinosporamide molecule [125]. Additionally, fluorinases have been used to make building blocks for the synthesis of other fluorinated molecules. Hong et al showed the biological production fluorinated polyketides. These molecules were generated from FA after the *in vivo* conversion to fluoroacetyl-CoA and subsequent consumption by minimal polyketide synthases [126].

1.6 Problem identification

D. cymosum has been shown to accumulate FA, presumably as a defence mechanism. However, research into this plant came to a halt when *S. cattleya* was also found to produce fluorinated metabolites, and the fluorination mechanism was never elucidated in higher plants. Following the identification of the fluorinase from *S. cattleya*, new information became available such as the gene sequence and natural substrate of bacterial fluorinases. But these aspects have not been investigated in *D. cymosum* to probe for fluorinating enzymes. Furthermore, although extremely rare, fluorinases from bacterial sources are the only representatives of enzymes that can catalyse a C-F bond. These enzymes have vast potential in biocatalytic applications and, as such, are highly sought after.

1.7 Aims and objectives

Aim 1: Sequence, assemble and annotate the transcriptome of *Dichapetalum cymosum* to characterise the wound response and investigate the possible occurrence of fluorometabolite biosynthetic genes.

1.1 Perform wounding studies on *D. cymosum* leaf tissue
1.2 Extract total RNA
1.3 Sequence the transcriptome of the unwounded and wounded tissue
1.4 Assemble and annotate the transcriptome data
1.5 Perform differential expression analysis between the unwounded and wounded tissue

Aim 2: Construct a proteome for *Dichapetalum cymosum* to investigate the possibility of a unique fluorinating enzyme

2.1 Collect flower, leaf and seed tissue of *D. cymosum*
2.2 Extract total protein
2.3 Fractionate protein and assay for fluorinase activity
2.4 Resolve on SDS-PAGE
2.5 Perform in-gel protein digestion and proteomic analysis using LC-MS/MS
2.6 Integrate with transcriptome data and analyse the proteome

Aim 3: Overexpress and characterise the fluorinase from *Actinopolyspora mzabensis* for potential biocatalytic application

3.1 Clone the *Am-fluorinase* gene into the expression host, *Escherichia coli*
3.2 Overexpress the recombinant protein
3.3 Recover the *Am*-fluorinase from inclusion bodies and refold the enzyme
3.4 Purify the *Am*-fluorinase and assay for fluorination activity
3.5 Characterise the *Am*-fluorinase in terms of optimal conditions for activity and stability

Chapter 2

Functional characterisation of the transcriptome from the fluoroacetate-producing plant, *Dichapetalum cymosum*, in response to mechanical wounding

Selisha A. Sooklal, Phelelani Mpangase, Shaun Aaron, Robert H. Archer and Karl Rumbold

This chapter has been prepared for publication.
Scientific Reports

Functional characterisation of the transcriptome from the fluoroacetate-producing plant, *Dichapetalum cymosum*, in response to mechanical wounding

Selisha A. Sooklal[1,+] , Phelelani Mpangase[2,+] , Shaun Aaron[2] , Robert H. Archer[3] & Karl Rumbold[1,*]

[1]Department of Microbiology, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, 2000, South Africa

[2]Sydney Brenner Institute for Molecular Biosciences, University of the Witwatersrand, Johannesburg, 2000, South Africa

[3]National Herbarium, South African National Biodiversity Institute, Pretoria, 0186, South Africa

[*]Corresponding author e-mail: karl.rumbold@wits.ac.za

[+]These authors contributed equally to this work

Abstract

The first fluorinated metabolite was isolated from the plant *Dichapetalum cymosum*, presumably as a defence mechanism. Given the rarity of fluorinated metabolites in nature, the biosynthetic origin of fluoroacetate in *D. cymosum* was of particular interest. However, the mechanism for fluorination was never elucidated. Here, we report on the first comprehensive transcriptome for *D. cymosum*. Mechanical wounding studies were performed and libraries of the unwounded (control) and wounded (30 and 60 minutes post wounding) plant were sequenced using the Illumina HiSeq platform. A combined reference assembly generated 77,845 transcripts. Using the SwissProt, TrEMBL, GO, eggNOG, KEGG, Pfam, EC and PlantTFDB databases, a 69% annotation rate was achieved. Differential expression analysis revealed the regulation of 364 genes in response to wounding. The wound responses in *D. cymosum* included key mechanisms relating to signalling cascades, phytohormone regulation, transcription factors and defence-related secondary metabolites. However, the role of fluoroacetate in wound responses remains unclear. Bacterial fluorinases were searched against the *D. cymosum* transcriptome but no transcript with homology was detected. This may suggest the presence of a different fluorinating enzyme in plants. The transcriptome produced in this study significantly increases genetic resources available for *D. cymosum*.

Introduction

Fluorinated metabolites are extremely rare in nature relative to other halogenated natural products[1,2]. This is largely a consequence of the low bioavailability of fluorine, the high energy cost associated with desolvating the fluoride ion and exclusion from enzymatic incorporation via halogenases that utilise oxidative catalytic mechanisms[2,3]. Despite these factors, a fluorinated secondary metabolite was isolated from the plant *Dichapetalum cymosum*[4,5].

D. cymosum, or more commonly known as gifblaar (Afrikaans for poison leaf), is a member of the *Dichapetalaceae* family characterised by its shrub-like appearance and extensive underground stem system[6]. This highly toxic plant is scattered in South Africa, Namibia, Zimbabwe, Botswana and Angola[7]. After observing sudden death in animals that ingested *D. cymosum*, Marais isolated fluoroacetate as the constituent responsible for toxicity[4,5].

More importantly, this marked the discovery of the first fluorinated metabolite from a biological source. *D. cymosum* was subsequently shown to accumulate fluoroacetate at concentrations up to 232 $mg \cdot kg^{-1}$ in the young leaves, 97 $mg \cdot kg^{-1}$ in the older leaves, 164 $mg \cdot kg^{-1}$ in the immature seeds and 362 $mg \cdot kg^{-1}$ in the flowers[8]. Toxicity is effectuated through 'lethal synthesis' whereby fluoroacetate is metabolised to (2*R*, 3*R*)-2-fluorocitrate *in vivo*. Coincidently, this is the only stereoisomer of fluorocitrate that is an inhibitor of aconitate hydratase, a key enzyme in the tricarboxylic acid cycle, thus terminating cellular respiration[9,10]. The presence of fluoroacetate has been confirmed in just over 30 plants worldwide, including 8 species from the *Dichapetalaceae* family[11].

The production of fluoroacetate is widely believed to be a defence mechanism in plants[12]. But, given how rare fluorometabolites are in nature, there was far more interest in the mechanisms underlying the biosynthesis of this fluorinated metabolite. Explanations such as the acquisition of fluoroacetate from microorganisms in the soil surrounding *D. cymosum* and fluoroacetate-producing endosymbionts have been proposed[13,14]. However, later work by Grobbelaar and Meyer showed that a callus of *D. cymosum*, devoid of contaminating microorganisms, had the ability to produce fluoroacetate when supplemented with fluoride ions[15]. These findings suggested the evolution of an enzyme with the capacity to incorporate fluorine into organic molecules. However, further attempts to clarify the mechanism for fluorination in *D. cymosum* were challenged by access to plant material and a general lack of knowledge regarding the natural substrate for such an enzyme[15-17]. Therefore, very little progress was achieved and the production of fluoroacetate remains a mystery in higher plants.

Attention quickly shifted to another fluorometabolite-producer, *Streptomyces cattleya*. This bacterium, which produces fluoroacetate as well as 4-fluorothrenine, provided a practical biological system to study fluorination[18]. Extensive efforts led to the discovery of the enzyme 5'-fluoro-5'-deoxyadenosine synthase (fluorinase) and elucidation of the fluorometabolite biosynthetic pathway in prokaryotes[2,19,20]. Fluorinases mediate the conversion of *S*-adenosyl-$_L$-methionine (SAM) and inorganic fluoride ions into 5'-fluoro-5'-deoxyadenosine (5'-FDA) and $_L$-methionine. This constitutes the first step in fluorometabolite biosynthesis. The pathway then proceeds with five additional enzymatic steps to yield fluoroacetate or 4-fluorothreonine. With the exclusive ability to catalyse the formation of a C-F bond, it is no surprise that fluorinases have significant potential in biotechnological applications[2].

Previous studies into *D. cymosum* were largely conducted on a physiological level. The progression of modern technologies such as next-generation sequencing (NGS) has allowed for the efficient profiling of non-model organisms, like *D. cymosum*, and has become the standard tool for simultaneous gene discovery and RNA quantification[21]. But there remains a severe lack in genetic resources available for this plant with only one sequence (KF147466.1) on the GenBank database. Furthermore, although the production of fluoroacetate is believed to be a defence mechanism, there are currently no reports detailing the wound response of *D. cymosum*. Wound responses such as the elicitation of signalling cascades, regulation of phytohormones and liberation of secondary metabolites have been well characterised in model species like *Arabidopsis thaliana*; however, very little is known about these responses in fluoroacetate-producing plants[22,23].

Herein, we explored the transcriptome of *D. cymosum* in response to mechanical wounding. Libraries were generated prior to mechanical wounding (T:0), as well as 30 minutes (T:30) and 60 minutes (T:60) post wounding. To the best of our knowledge, this is the first transcriptome of *D. cymosum*. Therefore, a combined reference transcriptome was created and comprehensively characterised, significantly expanding genetic resources available for this plant. In addition, the regulation of key processes in response to wounding was investigated, providing insights into the molecular mechanisms governing the defence response in *D. cymosum*.

Materials and methods

Materials

The cetyltrimethylammonium bromide (CTAB), polyvinylpyrrolidone (PVP), diethyl pyrocarbonate (DEPC), ethylenediaminetetraacetic acid (EDTA) and HCl were purchased from Sigma-Aldrich (Massachusetts, USA). The tris, NaCl, β-mercaptoethanol, chloroform, isopropanol, ethanol and tri-sodium citrate were from VWR (Pennsylvania, USA). The RiboLock was purchased from Thermo Fisher Scientific (Massachusetts, USA). Agarose gel electrophoresis was performed using the Mini-Sub cell GT electrophoresis system from BioRad (California, USA).

Collection of plant material and mechanical wounding

Dichapetalum cymosum (gifblaar) samples were collected at the Pretoria National Botanical Gardens, South Africa, in September 2015 during the spring season. In order to assess the genes associated with a wound response, mechanical wounding studies were performed according to[22]. An initial sample of the unwounded leaf tissue was taken at time zero (T:0) representing the control. Leaves, still attached to the plant, were then wounded. Approximately 40% of the leaf surface was punctured using sterile forceps. Leaves were severed from the plant at 30 minutes (T:30) and 60 minutes (T:60) post wounding. Technical replicates of samples at each time point were immediately frozen in liquid nitrogen and later transferred to a -80 °C freezer until further processing.

RNA isolation, construction of cDNA libraries and Illumina sequencing

Total RNA was isolated based on the method outlined by Xu et al[24]. In brief, leaf tissue was ground to a fine powder in liquid nitrogen using a pre-chilled mortar and pestle. The powder (~ 100 mg) was transferred to 0.6 ml of pre-warmed extraction buffer (100 mM tris-HCl, pH 8.0, 25 mM EDTA, 2 M NaCl, 2% CTAB, 2% PVP and 2% β-mercaptoethanol) and incubated at 65 °C for 15 min with occasional shaking. Chloroform (0.5 ml) was added to the slurry, mixed and then centrifuged. Following transfer of the upper aqueous phase to a new tube, 0.1 ml of 5 M NaCl and 0.3 ml of chloroform was added, mixed and centrifuged. This step was repeated and the upper aqueous phase was transferred to a new tube to which a half volume of isopropanol and a half volume of a high salt solution (0.8 M tri-sodium citrate dihydrate, 1.2 M NaCl) were added. The solution was incubated at room temperature for 15 min and the RNA was recovered by centrifugation. The pellet was washed with 75% ethanol, air dried and resuspended in DEPC-treated water supplemented with 1 µl of the RNase inhibitor, RiboLock. All centrifugation steps were carried out at 12,000 x *g* for 10 min at 4 °C. Total RNA was quantified using the Qubit Fluorometer 2.0 (Life Technologies) and RNA integrity was assessed using 1% agarose gel electrophoresis (Supplementary Table S1 and Supplementary Figure S1).

cDNA library construction and Illumina sequencing were performed at the Agricultural Research Council (Pretoria, South Africa) as a commercial service. Ribosomal RNA was depleted using the RiboZero Plant rRNA Removal Kit (Epicentre). Thereafter, cDNA

libraries for each time point (in triplicate) were created individually using the TruSeq Stranded mRNA Library Preparation Kit (Illumina). Paired-end sequencing was performed individually for each of the time-point samples and their respective replicates on the HiSeq 2500 (Illumina) platform using the HiSeq Reagent Kit v4 (Illumina).

Quality control and *de novo* assembly using Trinity

The web links for all tools and databases used in this study are provided in Supplementary Table S2. The quality of each library, before and after trimming, was analysed using FastQC version 0.11.2[25]. All adaptor sequences and short sequences (< 40 bp) were trimmed using Trimmomatic (version 0.30)[26]. Only the paired reads whereby both the forward and reverse read survived the processing were used for downstream analysis. Prior to the transcriptome assembly, all forward reads across all time points (T:0, T:30 and T:60) and replicates were combined into one file. The same was done for the reverse reads into a separate file to create a single RNA-seq data set. Using the combined forward and reverse reads, a comprehensive reference transcriptome (T:0 + T:30 + T:60) was assembled *de novo* using Trinity (version 2.0.2) under default settings[27].

Annotation of the *D. cymosum* transcriptome

In order to assign functional descriptions, Trinity transcripts were queried against the UniProt (SwissProt and TrEMBL), Gene Ontology (GO) and the evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG) databases using a blastx search with an e-value cut-off of 1E-05[28-30]. For GO, a set of plant-specific GO Slims were used to classify the annotations into 97 terms under the molecular function, cellular component and biological process categories. Annotations from eggNOG were allocated single-letter codes based on Clusters of Orthologous Groups (COG) to categorise descriptions into 25 functional groups. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used to map pathways represented in the *D. cymosum* transcriptome. This was performed using the web server, KOBAS 3.0[31].

Simple Sequence Repeats (SSR's) were identified in assembled transcripts using the Microsatellite (MISA) identification tool[32]. Search parameters included di-, tri-, tetra-, penta-, and hexa-nucleotide sequences that were repeated a minimum of ten, seven, five, five and five times respectively. The maximum number of bases allowed interrupting 2 SSR's was set at 100 bp.

TransDecoder (version 2.0.1) was then used to identify open reading frames (ORFs) with a minimum length of 100 consecutive amino acids from the assembled transcripts. HMMER (version 3.1b1) was used to scan the protein predictions for the presence of conserved domains against the Pfam database[33]. Additionally, protein sequences were queried against the ExPASy-Enzyme database to assign EC numbers[34], and against the Plant Transcription Factor Database (PlantTFDB) to identify transcription factors[35]. The e-value cut-off for blastp searches were 1E-05.

Investigating the *D. cymosum* transcriptome for putative fluorometabolite biosynthetic genes

The translated *D. cymosum* transcriptome, in all six frames, was searched for putative enzymes involved in fluorometabolite biosynthesis using blastp. Enzymes involved in fluorometabolite production in *S. cattleya* were obtained from GenBank and used as the query sequences. This included the fluorinase (WP_014144878.1), a purine nucleoside phosphorylase (PNP) (WP_014144879.1), an isomerase (WP_014142773.1), two aldolases (WP_014141855.1, WP_014150905.1), an aldehyde dehydrogenase (WP_014141713.1) and a 4-fluorothreonine transaldolase (WP_014151017.1). The best hit in the *D. cymosum* transcriptome for each of the queried enzymes was extracted and searched against the non-redundant (nr) database using blastp to obtain putative functions.

Abundance estimation and differential expression analysis

Abundance estimation and differential expression analysis were performed according to Haas et al using software included in the Trinity (version 2.0.2) package[27]. In order to assess differentially expressed genes (DEGs) in response to wounding, the high-quality clean reads from each library within time points T:0, T:30 and T:60 were individually

aligned to the assembled reference transcriptome (described in 'Quality control and *de novo* assembly using Trinity') using Bowtie. The relative abundance of each transcript was estimated using RNA-Seq by Expectation Maximization (RSEM) and differential expression analysis was conducted using edgeR. Expression values were normalised using the trimmed mean of *M*-values (TMM) method to account for inherent biases introduced through variations in sample composition and sequencing depth. DEGs with normalised expression values that adhered to a False Discovery Rate (FDR) of ≤ 0.001 and showed a $\geq$ 4-fold change between at least two time points were selected. This allowed for reliable estimations of relative levels which can be used to directly compare differences in gene expression between samples.

Annotation of DEGs

DEGs were searched against the SwissProt database and assigned GO terms using Function Annotator[36]. REVIGO was used to enrich GO processes associated with DEGs[37]. KEGG pathways were assigned using KOBAS version 3.0[31] and the PlantTFDB was used to identify transcription factors associated with the regulation of a wound response[35].

Results

Sequencing and *de novo* assembly of the *D. cymosum* transcriptome

In this study, *D. cymosum* leaves were challenged with mechanical wounding. cDNA libraries of replicates obtained from the control (T:0) and time points post wounding (T:30 and T:60) were individually sequenced using the Illumina HiSeq 2500 platform. At present, there is virtually no sequence information available for this plant with the GenBank database containing only 1 entry. Therefore, to develop the first comprehensive overview of the *D. cymosum* transcriptome, the data from the three libraries and their respective replicates were pooled to produce 60,638,658 raw paired-end (PE) reads with an average length of 125 bp. Following the data filtering process, an enhancement in the average quality (≥ 30 Phred score) at each base position of the sequence reads was seen (Supplementary Figure S2) and a total of 52,819,612 (87%) high-quality PE reads with a sequence length of 50-125 bp remained. The reads were then assembled *de novo* using Trinity. Trinity produced 77,845 transcripts with an average length and N50 of 454 bp and 480 bp, respectively. In addition, a GC content of 41.8% was observed. A summary of the sequencing, quality filtering and Trinity outputs are shown in Table 1. The shortest transcript generated was 224 bp while the longest transcript had a length of 10,795 bp. The length distribution of transcripts is depicted in Supplementary Figure S3.

Table 1. Summary of the sequencing, data filtering and assembly output for the *D. cymosum* transcriptome

Sequencing output	
Number of raw reads	60,638,658
Read length (bp)	125
Quality filtering output	
Number of high-quality reads	52,819,612
Read length (bp)	50-125
Number of PE reads used in assembly	26,409,806
Trinity output	
Total Trinity genes	64,325
Total Trinity transcripts	77,845
Average contig length (bp)	454
N50 (bp)	480
Percent GC	41.8

Functional characterisation of the *D. cymosum* transcriptome

Annotation overview

Sequencing of plant transcriptomes typically results in large amounts of deduced gene or protein sequences, of which, very few will ever be investigated experimentally. Annotation through comparative sequence analysis provides the only feasible route to assign putative functions[21]. Therefore, sequence similarity searches were conducted against various databases (SwissProt, TrEMBL, GO, eggNOG/COG, KEGG, Pfam, ExPASy-Enzyme and PlantTFDB). Collectively, a 69% annotation rate was achieved for the *D. cymosum* transcriptome using the aforementioned databases (Table 2) with majority of the transcripts being identified by TrEMBL. The organisms from which annotations were derived, was subsequently explored. The species distribution (Figure 1) showed that most sequences exhibited homology to *Arabidopsis thaliana* (27%). This was followed by *Populus trichocarpa*, (12.3%), *Jatropha curcas* (10.6%), and *Ricinus communis* (10.5%).

Table 2. Summary of the annotation output for the *D. cymosum* transcriptome

Database	**Number of annotated transcripts**
SwissProt	36,053
TrEMBL	53,452
GO	42,664
eggNOG	19,085
COG	18,690
KEGG	10,138 transcripts, 125 pathways
Database	**Number of annotated proteins**
Pfam	23,878
ExPASy-Enzyme	7,311
PlantTFDB	1,220

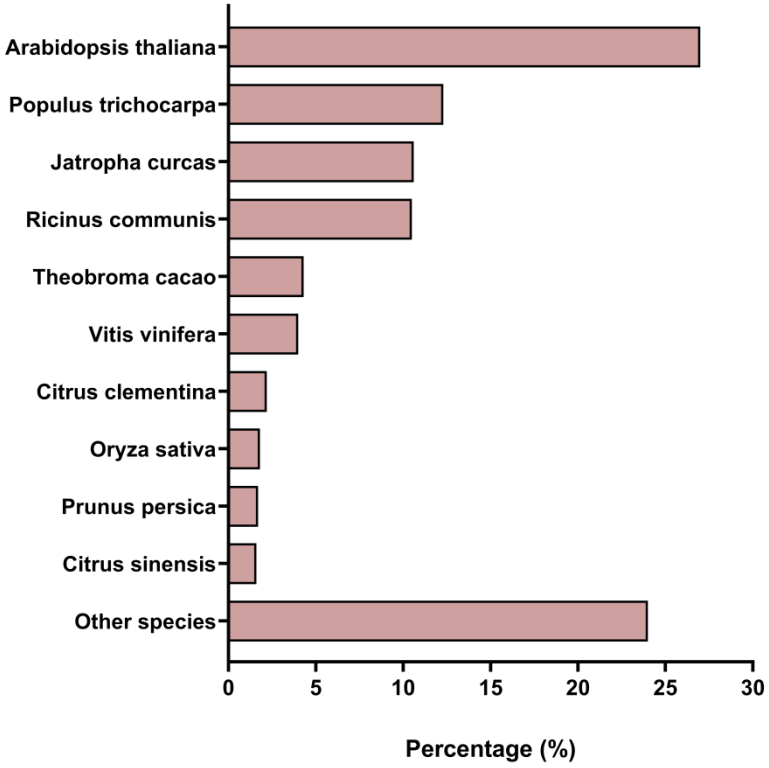


Figure 1. Species distribution of annotations obtained from the *D. cymosum* transcriptome. Majority of the *D. cymosum* transcripts annotated by the UniProt databases exhibited homology to sequences from *Arabidopsis thaliana*.

GO and eggNOG

Overall, 42,664 (58%) transcripts were assigned GO terms. In many cases, multiple GO terms were assigned to the same transcript thus resulting in the molecular function, cellular component and biological process domains containing 35,709; 31,638 and 31,136 transcripts, respectively. Using a set of plant-specific GO Slims, annotations were further sub-divided into 97 functional terms (Supplementary Figure S4). Amongst the molecular function category; ‘nucleotide binding’ and ‘binding’ were the most represented groups. Within the cellular component category, a high percentage of transcripts were assigned to ‘membrane’. For biological process, ‘cellular process’ and ‘metabolic process’ were the most dominant groups. Interestingly, it was observed that ‘response to stress’ and ‘signal transduction’ were also among the top represented groups in biological process. Since majority of the reads in this transcriptome were from the wounded *D. cymosum* leaves (T:30

+ T:60), it is not atypical to see a number of transcripts directed towards stress and recovery responses of the plant.

Orthologous inferences were made using eggNOG. Here, 19,085 transcripts were identified, of which, 18,690 received single-letter codes as defined for COG to allow functional clustering into fewer groups (Figure 2). 'Signal transduction mechanisms' was the most abundant group, again, suggesting the common response of a wounded plant. This was followed by 'general function prediction only'.

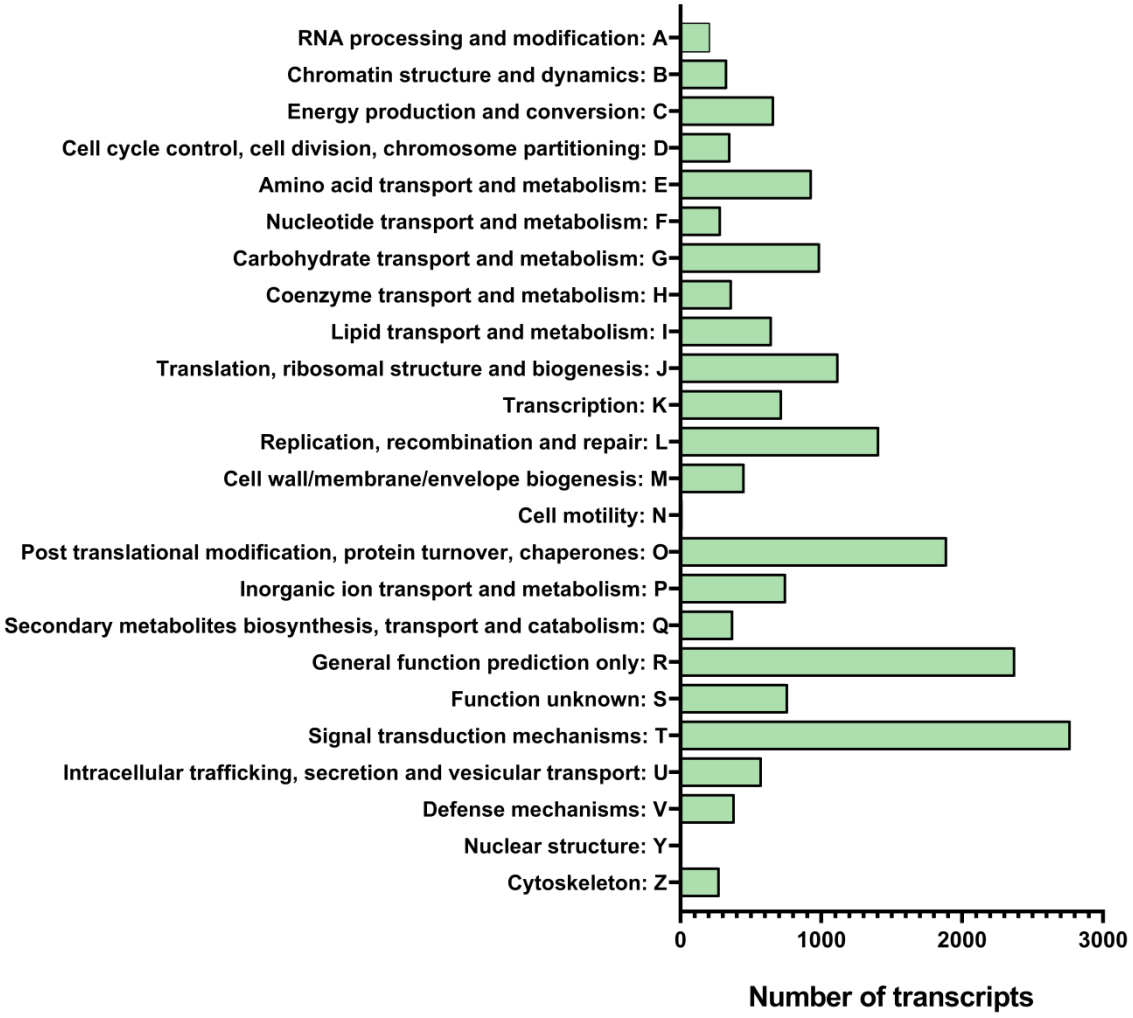


Figure 2. COG classification of *D. cymosum* transcripts. A total of 19,085 transcripts were identified by the eggNOG database. Of these, 18,690 were categorised into 25 COG clusters.

KEGG

Active pathways in the *D. cymosum* transcriptome were identified using *A. thaliana* as a reference on the KEGG database. KEGG successfully mapped 10,138 individual transcripts to 125 unique pathways within the 'metabolism', 'genetic information processing', 'environmental information processing', 'cellular processes' and 'organismal systems' categories (Figure 3A). Pathways within the 'metabolism' category accounted for 55% of the aforementioned transcripts. Of these transcripts, nearly 10% were allocated to selected secondary metabolic processes, as shown in Figure 3B. The 'terpenoid backbone biosynthesis', phenylpropanoid biosynthesis' and 'ubiquinone and other terpenoid-quinone biosynthesis' secondary metabolic pathways were highly represented.

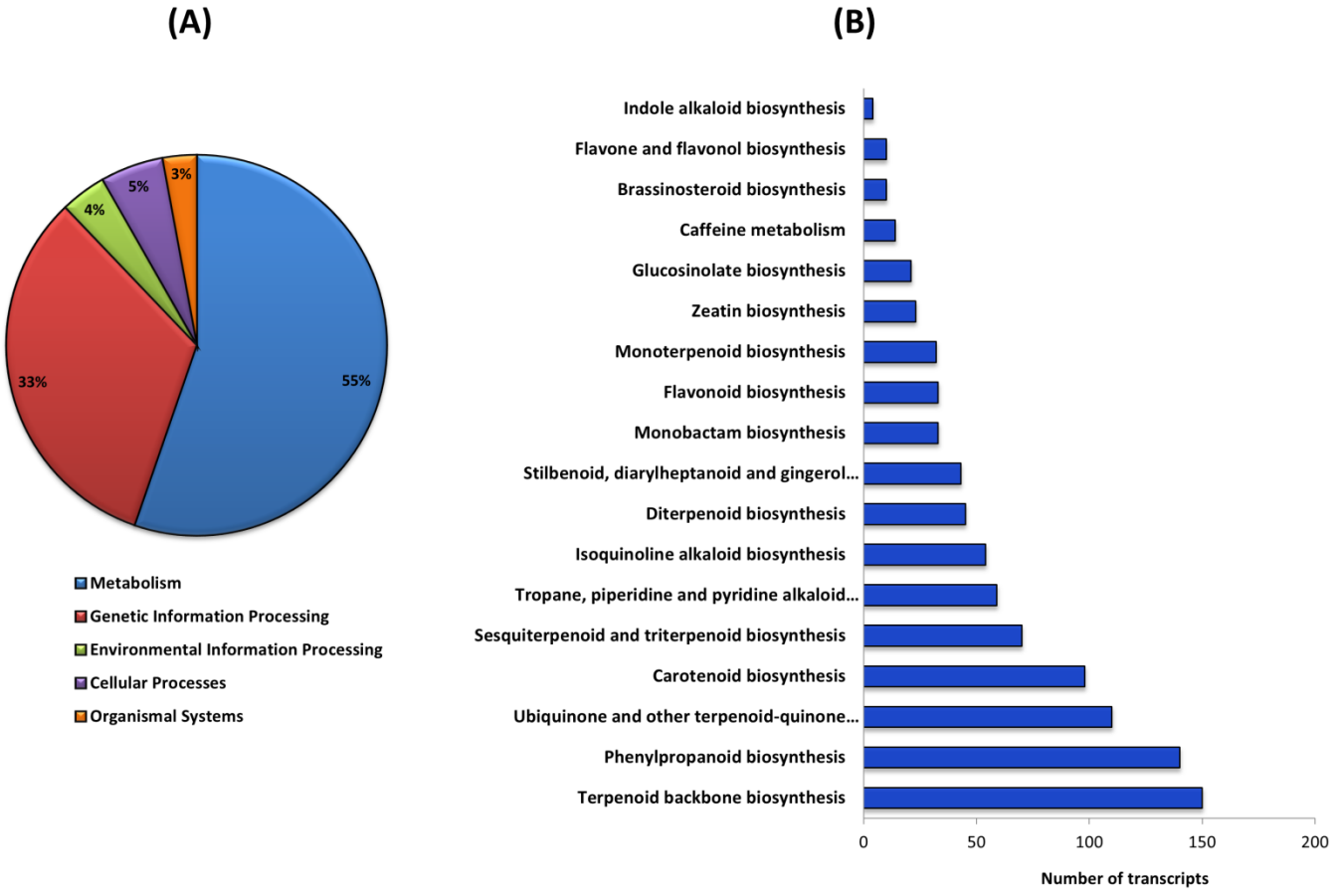


Figure 3. (A) Distribution of 10,138 transcripts into the broader metabolism, genetic information, environmental information processing, cellular processes and organismal systems categories in KEGG. (B) Allocation of transcripts within 'metabolism' into selected secondary metabolic pathways.

SSR's

In total, 1,534 SSR's were discovered in 1,442 *D. cymosum* transcript sequences. Of these, 91 sequences contained more than one SSR. The trinucleotide motifs accounted for 57.8% (887) of all SSR's. This was followed by dinucleotide motifs with 25.6% (393) of SSR's (Supplementary Table S3). The most abundant repeat sequence for the trinucleotide motifs were CTT/AAG and AG/CT for the dinucleotide motifs. In addition, 51 SSR's were detected in a complex formation.

Pfam and EC

Overall, 36,528 ORFs of at least 100 amino acids long were identified. The predicted protein sequences were queried against the Pfam database to extract conserved domain information. Of these, 23,878 (65.4%) sequences contained at least one Pfam domain. Collectively, 8,496 different Pfam domains were ultimately assigned with the 'protein kinase' and 'protein tyrosine kinase' domains being the most predominant by far. In addition, 976 unique EC numbers were retrieved from ExPasy for 7,311 protein sequences. EC numerically classified enzymes into 6 categories based on catalytic function (Supplementary Figure S5). Specifically, these were transferases (45%), hydrolases (27%), oxidoreductases (14%), lyases (5%), ligases (5%) and isomerases (4%).

Transcription factors

In all, 1,220 transcription factors from 50 unique families were found in the *D. cymosum* transcriptome. The most number of transcription factors (99) were from the MYB-related family. This was followed closely by the NAC and WRKY families containing 96 and 92 transcription factors, respectively (Figure 4).

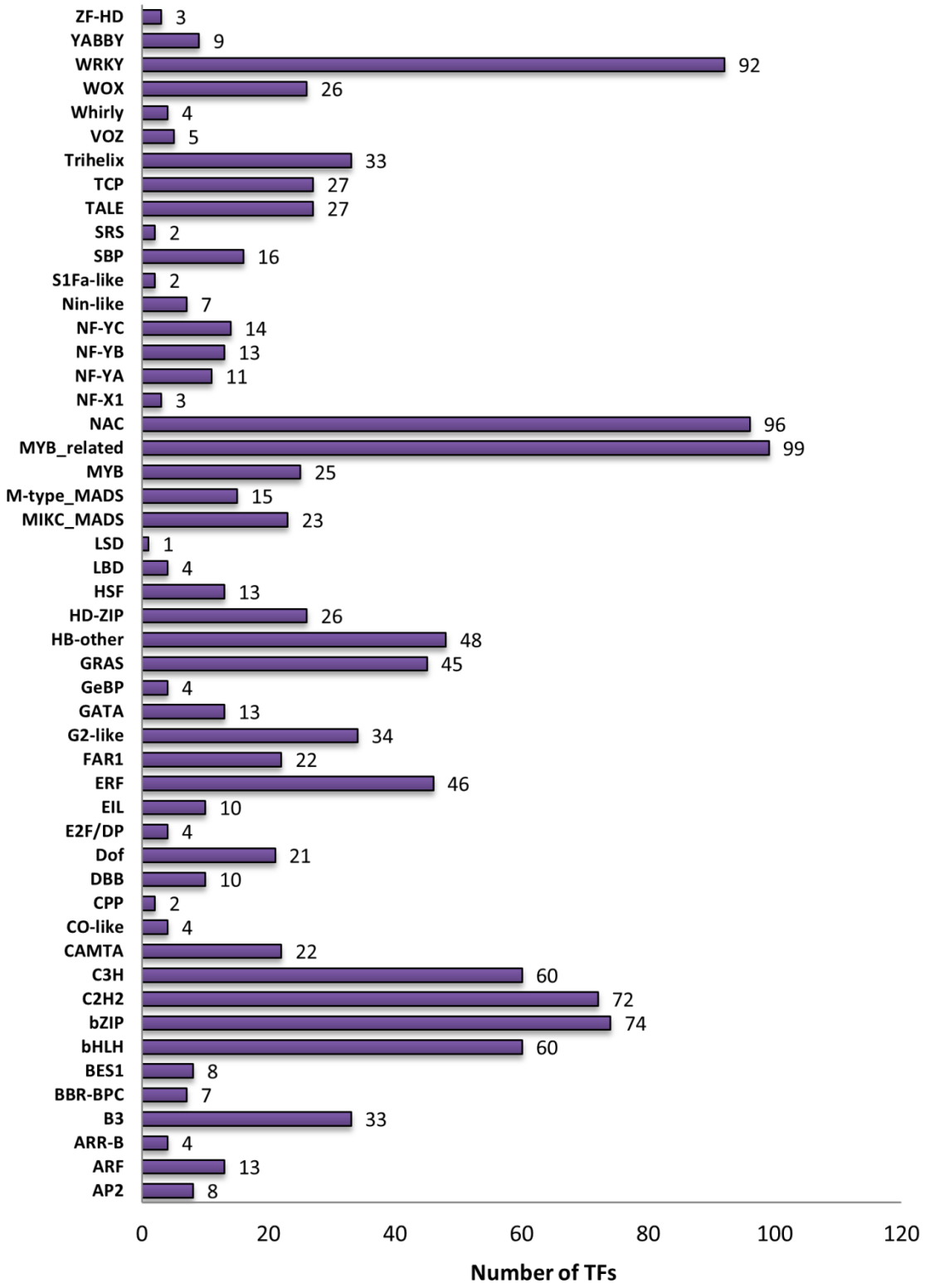


Figure 4. The families of transcription factors detected in the *D. cymosum* transcriptome. A total of 1,220 transcription factors were assigned to 50 unique families using PlantTFDB.

Investigating the *D. cymosum* transcriptome for fluorometabolite biosynthetic genes

Following extensive annotation of the *D. cymosum* transcriptome, only four putative fluoride ion transporters were detected (TR45147_c0_g1_i1, TR45147_c0_g1_i2, TR45147_c0_g1_i3 and TR45147_c0_g1_i4) and transcripts associated with fluorinase activity were not found. Consequently, the enzymes involved in the *S. cattleya* fluorometabolite biosynthesis pathway were queried against the *D. cymosum* transcriptome in order to identify any potential hits that may have been annotated fallaciously or not annotated at all. The best hit for each enzyme is summarised in Supplementary Table S4. Hits in the *D. cymosum* transcriptome exhibited low identity to the enzymes from *S. cattleya*. Moreover, the matches had high e-values, low bit scores and were further complicated by short alignment lengths suggesting that no significant match was found, except for an aldolase (56% identity) and an aldehyde dehydrogenase (41.5% identity).

Characterisation of the wound response in *D. cymosum*

Until now, there have been no reports detailing the metabolic adjustments of *D. cymosum* in response to wounding. Hence the genetic changes in *D. cymosum* at two stages following mechanical wounding were characterised by aligning the reads from each time point (T:0, T:30 and T:60), and their respective replicates, to the reference transcriptome generated in this study. The correlation matrix showed minor variation in expression between replicates within the same time point indicating reproducibility of the experimental data (Figure S6). However, as expected, more variation was seen across time points between the healthy and wounded *D. cymosum* leaves. The abundance of each transcript was then estimated and differential expression analysis was conducted using edgeR. Search parameters (FDR $\leq$ 0.001 and $\geq$ 4-fold change between two time points) limited DEGs to only those that were the most significantly differentially regulated and, hence, are pivotal in the wound response. Using this approach, 364 DEGs were identified (Figure 5A). Of these, 223 DEGs were transiently up-regulated in at least one time point post wounding and, conversely, 141 DEGs were transiently down-regulated in at least one time point post wounding (Figure 5B).

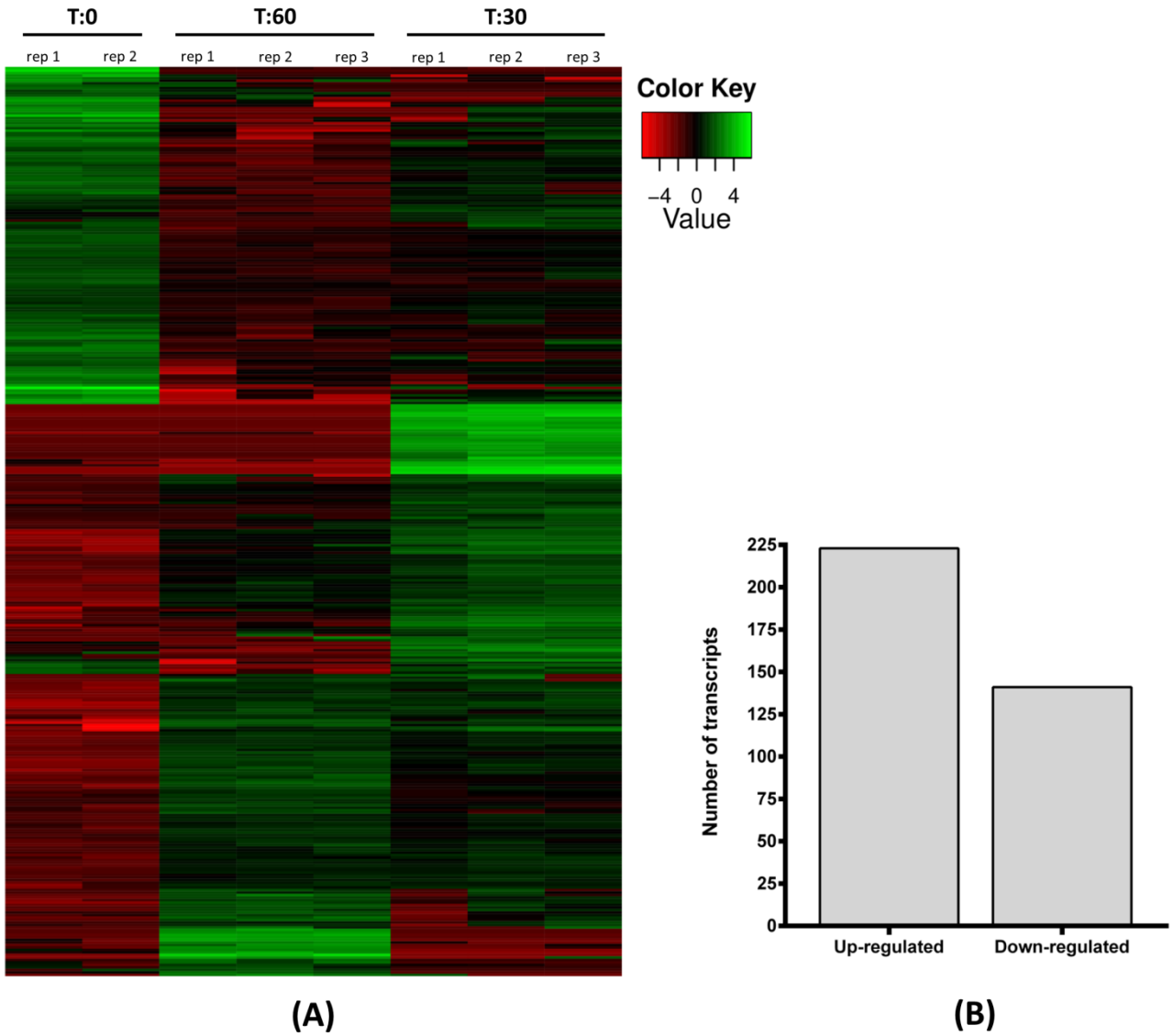


Figure 5. (A) Clustered heatmap showing the differential regulation (FDR ≤ 0.001) of 364 genes in response to mechanical wounding in *D. cymosum*. Green indicates up-regulation while red indicates down-regulation. (B) Of the 364 DEGs detected, 223 were up-regulated and 141 were down-regulated.

Enriched gene ontologies of DEGs

In total, 278 DEGs were annotated with 1,972 GO terms. The molecular function, cellular component and biological process categories contained 236, 213 and 247 transcripts, respectively. Enrichment of GO terms was then carried out using REVIGO. Enrichment allowed for functional profiling of DEGs by finding a subset of representative terms in order to elucidate underlying mechanisms associated with the wound response in *D. cymosum*. These processes are shown on a scatter plot (Figure 6). The log count of GO terms for each subcategory is represented on the y-axis as well as by the colour of the bubble. The x-axis is ordered based on semantic differences between GO terms. For molecular function, it is evident that 'signal transducer activity', 'transcription factor activity' and 'calcium ion binding' are among the most enriched processes. For biological process, 'response to ethylene' was most represented. Of note were also 'ethylene biosynthesis', 'jasmonic acid biosynthesis' and 'camalexin biosynthesis'. For the cellular component category, the 'endomembrane system', 'mitochondrion' and 'cell wall' were prioritised. It is important to note, however, that the plot is based on the frequency of the GO terms rather than representing the actual regulation of each gene. This means that if a subcategory appears lower down on the plot, there are fewer terms associated with this category rather than the process being down-regulated.

Regulation of DEGs

In comparison to the enriched GO plots, the absolute regulation of each gene was also investigated. For this, annotations for 247 DEGs were retrieved from the SwissProt database and transcripts were arranged based on putative function. Various *D. cymosum* genes associated with phytohormones, transcription factors, signal transduction mechanisms, redox regulation, proteolysis, secondary metabolite biosynthesis and pathogenesis-responsive pathways were found to be differentially regulated in response to mechanical wounding. Figure 7 and Supplementary Table S5 shows the number of DEGs associated with these responses that were either up-regulated or down-regulated in at least one time point post wounding.

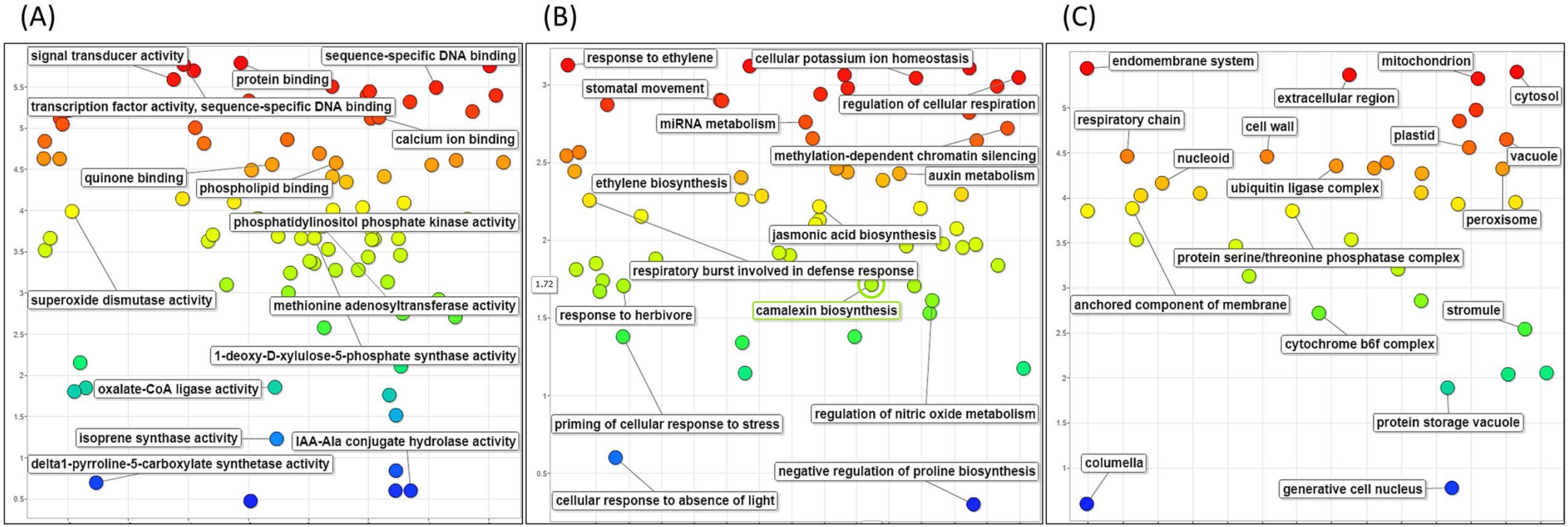


Figure 6. Scatter plot of the GO enrichment for DEGs in response to mechanical wounding in *D. cymosum*. Enriched DEGs were classified into the (A) molecular function (B) biological process and (C) cellular component categories. The plot is based on the frequency of the GO terms rather than representing the actual regulation of each gene. The log count is shown on the y-axis. Semantic space is represented on the x-axis.

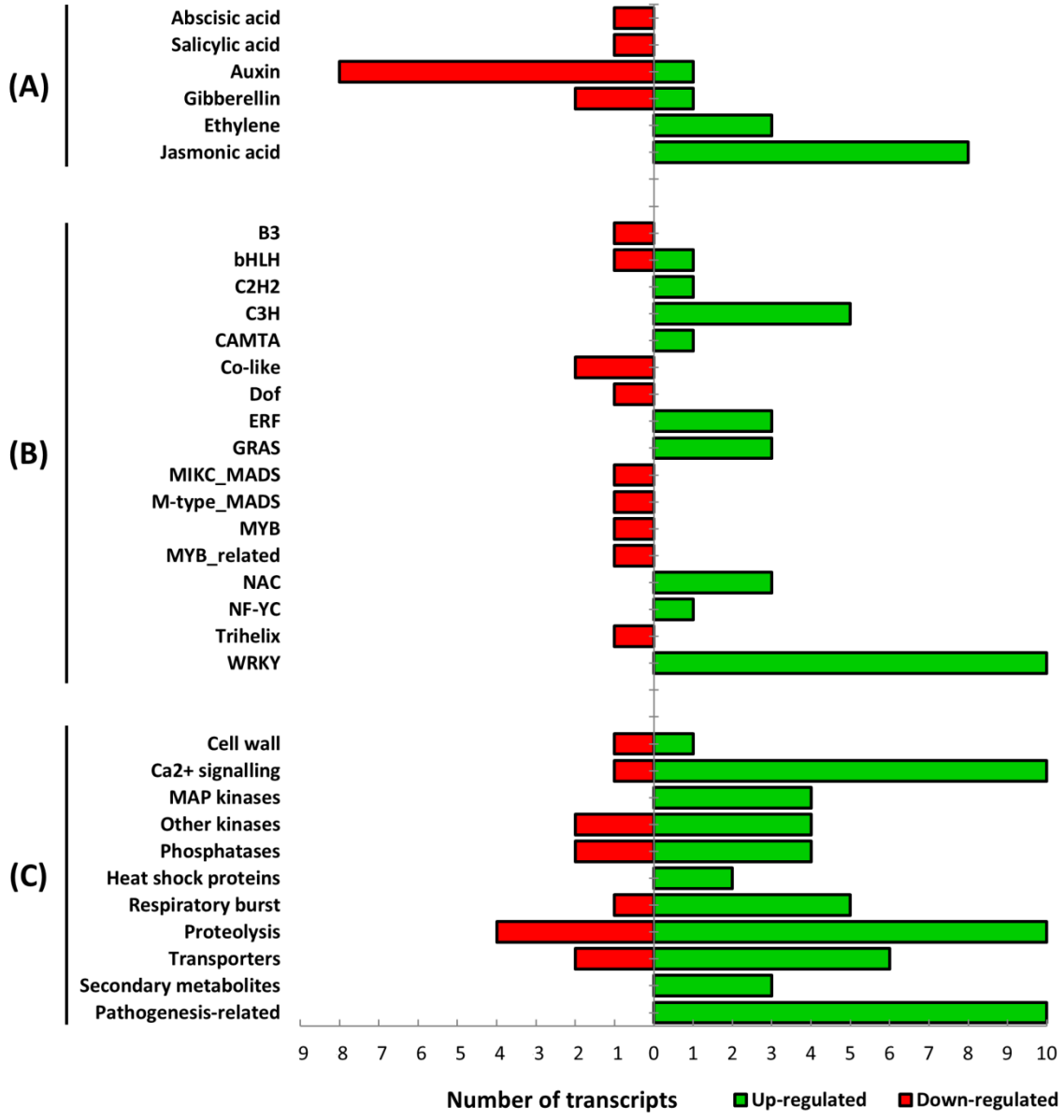


Figure 7. The number of DEGs either up-regulated (green) or down-regulated (red) in wound-responsive processes relating to (A) phytohormones, (B) transcription factors and (C) other significant processes in *D. cymosum*.

The analysis of phytohormones (Figure 7A, Supplementary Table S5) indicated a prominent up-regulation of the jasmonic acid (JA), ethylene (ET) and salicylic acid (SA) pathways following wounding, whereas the abscisic acid (ABA), auxin (IAA) and gibberellin (GA) pathways appeared to be down-regulated in *D. cymosum*. For the JA pathway, this included the persistent up-regulation of genes involved in JA biosynthesis such as linoleate 13S-lipoxygenase (*13-LOX*) and 12-oxophytodienoate reductase 3 (*OPR3*) at T:30 and T:60. In addition, several JASMONATE ZIM-DOMAIN (*JAZ*) homologues were up-regulated too. Genes such as *S*-adenosylmethionine synthase 2 (*SAM2*) and 1-aminocyclopropane-1-carboxylate oxidase (*ACO*) involved in the biosynthesis of ET were rapidly increased (T:30 + T:60) in response to wounding. Consistently, ethylene-responsive transcription factors (ERFs) followed the same regulation. The enhanced disease resistance 2 gene (*EDR2*), a negative regulator of SA-mediated resistance[38], was expressed at both T:0 and T:30. However, a down-regulation at T:60 was seen suggesting SA activity at a later stage post wounding.

Conversely, the AUX pathway was down-regulated following mechanical wounding. A number of transcripts corresponding to tryptophan aminotransferase-related proteins (*TAR1*), involved in AUX biosynthesis, as well as auxin-responsive proteins decreased in expression levels by T:60. Reduction in expression of the abscisic stress-ripening protein 3 as well as of the gibberellin-2-beta-dioxygenase and gibberellin receptor GID1 by T:60 also indicated the down-regulation of the ABA and GA pathways, respectively.

The regulation of transcription factors constituted the largest group of DEGs. A total of 38 putative transcription factors (28 up-regulated, 10 down-regulated upon wounding) were detected within 17 families (Figure 7B, Supplementary Table S5). Notably, transcription factors from the WRKY family (WRKY6, WRKY15, WRKY18, WRKY33, WRKY46 and WRKY53) were most prominently up-regulated at both T:30 and T:60. This was followed by the C3H, ERF and NAC families with consistent up-regulation following wounding. Interestingly, transcription factor families such as the Co-like, MYB, MYB-related and trihelix were systematically down-regulated.

Figure 7C and Supplementary Table S5 presents other processes vital in the wound response of *D. cymosum*. Signal transduction mechanisms included the Ca^{2+} signalling pathway, the production of reactive oxygen species (ROS) and phosphorylation-dephosphorylation events carried out by mitogen-activated protein kinases (MAPK), other

kinases and phosphatases. Within the Ca^{2+} signalling pathway, several calcium-transporting ATPases, calmodulins (*CaM1*), calmodulin-like proteins (*CML27* and *CML45*) and calmodulin-binding transcription activators were rapidly up-regulated at both T:30 and T:60. In stark contrast, a calcineurin B-like protein 4 (*CBL4*) was down-regulated. The up-regulation (T:30 and T:60) of a catalase, reticuline oxidase, thioredoxin and glutaredoxin suggested a respiratory burst following wounding. Expression levels of all MAP kinases (including MAPKKK1 and MAPK3) were increased at T:30 as well as T:60, however, other kinases exhibited a mixed regulation with four up-regulated and two down-regulated following wounding. Similarly, of the six protein phosphatases detected, four were up-regulated and two were down-regulated upon wounding. Several disease resistance proteins consistent with pathogenesis-related proteins were also found up-regulated from T:30. Furthermore, the production of heat shock proteins and selected secondary metabolites were induced at a later stage following wounding. Both 1-deoxyxylulose-5-phosphate synthase (*DXPS*) and isoprene synthases (*ISPS*), involved in terpenoid biosynthesis, were up-regulated at T:60 only.

Discussion

D. cymosum is the first plant in which fluoroacetate, a fluorinated organic metabolite, was observed[5]. Given the rarity of the C-F bond in nature, *D. cymosum* represents a potentially significant source for fluorinating enzymes. However, tangible research into *D. cymosum* is outdated and prior attempts to elucidate the fluorination mechanism have been unsuccessful. Genetic information offers a practical alternative to study biological mechanisms in plants[39,40]. But, a significant limitation has been the lack of sequence data publically available for *D. cymosum*. In this study, RNA-seq was used to produce the first overview of the transcriptome for *D. cymosum*, as well as to increase our understanding of the wound-response in this plant at two stages following mechanical wounding.

Sequencing was performed using the Illumina platform and reads obtained from the unwounded (T:0) and wounded (T:30 and T:60) leaves were combined in order to generate a comprehensive transcriptome, representing a broad diversity of genes, for *D. cymosum*. This assembly produced 77,845 transcripts. With the primary goal of identifying a potential fluorinase gene, the assembled transcriptome was extensively characterised using various

databases. In doing so, a 69% annotation rate was achieved, bordering the upper limit of that expected for plants without a reference genome[21]. However, a transcript equated with fluorination activity was not detected. Therefore, using the bacterial fluorometabolite biosynthetic genes, we searched the *D. cymosum* transcriptome to find transcripts with homology that were not previously identified. Irrespectively, no significant hit was obtained, except for an aldolase and an aldehyde dehydrogenase, and the remaining matches most likely arose from remote sites of homology.

Fluorinases have the exclusive ability to catalyse a C-F bond and thereby facilitate the first step in fluorometabolite biosynthesis[19,41]. At present though, there are only five bacterial fluorinases and none with eukaryotic origin have been found[42-44]. Considering the dissimilarities between bacteria and higher plants, this not only leaves a gap in our knowledge regarding eukaryotic fluorinases, but inherently limits homology searches in higher plants. Zhu et al also reported that there had been no sequence match between bacterial fluorinases and genomes from the plant kingdom, and suggested the evolution of more than one biological mechanism for fluoride incorporation[45]. While it is possible that the sequencing depth of the *D. cymosum* transcriptome may have been insufficient to detect transcripts with low abundance, the failure to detect any fluorinases may also support the hypothesis of a distinctly different fluorinating enzyme in plants. Consequently, a novel gene for fluorination may be present in this transcriptome, however, remains unidentified.

In contrast to the rarity of fluorinases, aldolases and aldehyde dehydrogenases appear to be more ubiquitous in bacteria, fungi, algae, plants and animals. In *S. cattleya* alone, four putative aldolases were identified. The knockout mutation of one aldolase and *in vitro* purification of another excluded their involvement in the fluorometabolite pathway[46]. The two remaining aldolases were searched for in the *D. cymosum* transcriptome and one produced a hit with 56% identity. Given that the exact aldolase involved in fluorometabolite biosynthesis remains elusive[46], it is difficult to decipher if this transcript plays a role in the pathway. The aldehyde dehydrogenase from *S. cattleya*, on the other hand, produced a hit with 41.5% identity in the *D. cymosum* transcriptome. This enzyme was shown to accept substrates other than fluoroacetaldehyde and, moreover, Murphy et al showed that a yeast aldehyde dehydrogenase was also able to catalyse the conversion of fluoroacetaldehyde to fluoroacetate[47]. Aldolases and aldehyde dehydrogenases are not exclusive to the fluorometabolite pathway[48,49]; therefore it is not surprising that some homology was found with transcripts in the *D. cymosum* transcriptome. Further annotation of these transcripts

verified their putative function as an aldolase and aldehyde dehydrogenase. Nonetheless, their particular involvement in the fluorometabolite pathway would need to be confirmed, assuming the pathway follows the same course in plants.

While fluoroacetate is widely believed to constitute a defence mechanism, very little is understood about the underlying wound response in *D. cymosum*. Hence, we aimed to characterise the key responses following mechanical stress. Apart from gene discovery, RNA-seq has emerged as an efficient technology for quantifying relative transcript abundance as its accuracy rivals that of other well established methods, such as microarrays and quantitative polymerase chain reaction (qPCR)[21,50-52]. Nevertheless, to ensure reliable results without the need for further validation, only the most significant changes were reported in this study. Overall, 364 DEGs were found which predominantly corresponded to signal transduction mechanisms, phytohormone regulation, transcription factors and defence-related proteins; demonstrating certain hallmarks of the wound response in *D. cymosum*.

Some of the first molecular events upon mechanical wounding in *D. cymosum* were constituted by Ca^{2+}, ROS and phosphorylation-dephosphorylation signalling cascades. These molecules, in conjunction with phytohormones, form part of an interconnected signal transduction mechanism essential in regulating cellular responses upon abiotic stress[53]. In *A. thaliana*, mechanical wounding has been shown to trigger a rapid influx of calcium ions[54]. Calcium ions modulate downstream processes by either stimulating the production of ROS ($^1O_2$, O_2^-, OH^- and H_2O_2)[55] or through the action of calcium-binding proteins (CaMs, CMLs, CBLs and CDPKs)[53,56]. ROS-scavenging enzymes such as the catalase, reticuline oxidase, thioredoxin and glutaredoxin were steadily up-regulated in *D. cymosum*, suggestive of a respiratory burst and subsequent processing of ROS following mechanical stress. ROS has been linked to the activation of hormonal pathways and phosphorylation-dephosphorylation cascades. However, the unabated production of ROS can result in severe oxidative damage to DNA, RNA and proteins. As such, ROS-scavenging enzymes are produced to neutralise elevated levels[53,57]. Likewise, several CaMs and CMLs identified in this study were highly up-regulated as a result of wounding, while a CBL4 protein was down-regulated. CaMs and CMLs perceive stress signals by directly binding Ca^{2+} following the influx[56]. Previous studies have reported that conformational changes in CaMs and CMLs, induced upon Ca^{2+}-binding, affects interactions with a complex array of target proteins allowing for appropriate stress responses[58]. The down-regulation of a CBL protein correlates with other studies that

have implicated this protein in osmotic stress rather than mechanical stress[59,60]. CaMs and CMLs are also independently capable of activating phosphorylation-dephosphorylation cascades to elicit a defence response. Numerous kinases and phosphatases which modulate responses through phosphorylation patterns were found in the present study. However, among the most notable were MAP kinases. Both MAPKKK1 and MAPK3, associated with stress responses in *A. thaliana*[61,62], were persistently up-regulated in *D. cymosum*. Similar signalling responses have been observed in *A. thaliana*[63]. Moreover, Walley et al went on to show that these responses (Ca^{2+}, MAPK and WRKY) are highly up-regulated in *A. thaliana* just five minutes after mechanical wounding[63]. There is also evidence for extensive cross-talk between these signalling cascades and phytohormone responses[53].

Phytohormones regulate a variety of plant processes such as growth, development and defence or tolerance to abiotic and biotic stresses. The influence of these molecules on various stresses has been widely studied[64]. The findings in this study demonstrated that the JA, SA and ET pathways in *D. cymosum* were up-regulated in response to wounding while the IAA, GA and ABA pathways were down-regulated. The IAA and GA pathways are typically involved in plant development[65]. But upon wounding, there is a molecular reconfiguration to alter from a growth state to a defence state, therefore, it is not unusual to see a reduction in growth related hormones. The ABA pathway is associated with seed dormancy and stomatal closure, implicating this phytohormone in drought and osmotic stress responses[66]. In contrast, although also involved in developmental processes; the JA, SA and ET pathways have more frequently been affiliated with defence regulation in plants[65]. The JA pathway, which plays the most conspicuous role, is linked to tissue damage caused by mechanical wounding, insects, herbivores and necrotrophic pathogens[64]. The detection of both early (13-LOX) and late (OPR3) enzymes involved in the biosynthesis of JA in *D. cymosum* suggested the rapid response of JA to mechanical wounding. In addition, several JAZ proteins were also found to be up-regulated, most predominantly at T:60. Under normal conditions, JAZ proteins regulate JA-responses by acting as repressors of MYC transcription factors. However, following wounding, there is rapid synthesis of JA. Jasmonates have been shown to bind the receptor, CORONATINE INSENSITIVE 1 (COI1), complexed with JAZ. This targets JAZ for degradation and releases MYC allowing for transcription of wound-responsive genes. The interaction between JA and JAZ provides tight regulation of responses, however, the excessive amplification of JAZ proteins also suggested a decline in the JA-pathway at T:60[64]. Interestingly, this is when SA started to

show some activity in *D. cymosum*, evident by the down-regulation of EDR2. The antagonistic relationship between JA and SA has been reported and the findings here are coherent with other studies[67,68]. Similar to *D. cymosum*, Lee et al showed a rapid increase in JA levels 30 minutes after mechanical wounding in *Oryza sativa* (rice) leaves. This JA-burst also coincided with a decline in SA levels[68]. SA is typically activated in response to biotrophic pathogens and mediates pathogen resistance[66]. Given that mechanical wounding not only damages plant tissue, but also provides a potential entry point for pathogens, it is not surprising to see SA up-regulated at a later time point in *D. cymosum*. Again, detection of both early (SAM2) and late enzymes (ACO) involved in the biosynthesis of ET[69] in *D. cymosum* suggested activity of the pathway. However, in contrast to SA, ET is known to operate synergistically with JA to activate expression of defence genes[65,66]. Cross-talk between phytohormone signalling pathways (JA, SA and ET) is a mechanism that effectively allows for the regulation of stress responses. This regulation primarily occurs at the level of transcription[54].

Transcription factors constituted one of the largest groups of genes differentially expressed in *D. cymosum* following mechanical stress. Transcription factors act directly on stress-responsive genes, such as defence related genes or secondary metabolites, and are thus one of the most crucial aspects in mounting a tailored response[70]. In the present study, WRKY transcription factors were most prominently up-regulated. WRKY transcription factors are commonly associated with a variety of stress responses[71]. WRKY15, WRKY18, WRKY33 and WRKY46 particularly, have been associated with pathogen infection, drought stress, osmotic stress and mechanical wounding[72-78]. In addition, ERFs are the major target of ET signalling and act on a variety of defence related genes. Secondary metabolites from the phenylpropanoid and terpenoid pathways have been shown to play roles in stress responses[79-82]. Here, only the terpenoid pathway was found amongst the most differentially regulated suggesting an important role for terpenoid metabolism in the wound response in *D. cymosum*.

Conclusions

D. cymosum is a significant part of history representing the source of the first fluorinated metabolite. In the present study, a comprehensive transcriptome of *D. cymosum* was

produced. To the best of our knowledge, this is the first transcriptome from *Dichapetalaceae*, a family comprising of several FA-producing plants. The gene expression profiles of *D. cymosum* were evaluated in response to wounding. This revealed the fine-tuning of signalling cascades, phytohormone regulation, transcription factors and secondary metabolites most important in the wound response. However, the role of FA in wound responses remains unclear. The transcriptomic data produced in this study has been made available and will have a significant impact on future research into the origin of FA in plants.

Competing interests

The authors declare no competing interests.

Acknowledgements

We kindly thank Mihai-Silviu Tomescu for technical assistance with aspects of the annotation for the transcriptome. This work is based on the research supported in part by the National Research Foundation of South Africa (Grant No. 95062 to S. A. S.). The opinions, findings and conclusions or recommendations expressed in this publication are that of the author(s) and the NRF accepts no liability whatsoever in this regard.

Author contributions

S.A.S., R.H.A. and K.R. conceptualised the study. S.A.S performed the wounding studies and the RNA extractions. P.M. assembled the transcriptome and performed the differential expression analysis. S.A.S. and P.M. annotated the transcriptome. S.A. assisted with methodology for transcriptome analysis. S.A.S. analysed the data and wrote the manuscript. P.M., S.A., R.H.A. and K.R. reviewed the manuscript. K.R. supervised the project.

References

1. Gribble, G. W. (2003). The diversity of naturally produced organohalogens. *Chemosphere*, **52**, 289-297.
2. O'Hagan, D., & Deng, H. (2014). Enzymatic fluorination and biotechnological developments of the fluorinase. *Chemical Reviews*, **115**, 634-649.
3. Harper, D. B. and O'Hagan, D. (1994). The fluorinated natural products. *Natural Product Reports*, **11**, 123-133.
4. Marais, J. S. C. (1943). The isolation of the toxic principle "potassium cymonate" from "Gifblaar" *Dichapetalum cymosum* (Hook) Engl. *Onderstepoort Journal of Veterinary Science and Animal Industry*, **18**, 203-206.
5. Marais, J. S. C. (1944). Monofluoroacetic acid, the toxic principle of "Gifblaar", *Dichapetalum cymosum* (Hook) Engl. *Onderstepoort Journal of Veterinary Science and Animal Industry*, **20**, 67-73.
6. Nel, E., Robbertse, P. J., & Grobbelaar, N. (1982). A morphological study of *Dichapetalum cymosum* (Hook.) Engl. Part 1. The underground stem and root systems. *South African Journal of Botany*, **1**, 14-17.
7. De Sousa Correia, R. I., & Van Rensburg, L. (2000). Geographical distribution and local occurrence of *Dichapetalum cymosum* (Hook) Engl.(gifblaar) in Namibia and surrounding areas. *Agricola*, **11**, 19-28.
8. Meyer, J. M., & Grobbelaar, N. (1990). The determination, uptake and transport of fluoroacetate in *Dichapetalum cymosum*. *Journal of plant physiology*, **135**, 546-549.
9. Peters, R. A., Wakelin, R. W., Buffa, P., & Thomas, L. C. (1953). Biochemistry of fluoroacetate poisoning. The isolation and some properties of the fluorotricarboxylic acid inhibitor of citrate metabolism. *Proceedings of the Royal Society of London. Series B-Biological Sciences*, **140**, 497-506.
10. Treble, D. H., Lamport, D. T. A., & Peters, R. A. (1962). The inhibition of plant aconitate hydratase (aconitase) by fluorocitrate. *Biochemical Journal*, **85**, 113.
11. Lee, S. T., Cook, D., Pfister, J. A., Allen, J. G., Colegate, S. M., Riet-Correa, F., & Taylor, C. M. (2014). Monofluoroacetate-containing plants that are potentially toxic to livestock. *Journal of Agricultural and Food Chemistry*, **62**, 7345-7354.
12. Leong, L. E. X., Khan, S., Davis, C. K., Denman, S. E., & McSweeney, C. S. (2017). Fluoroacetate in plants-a review of its distribution, toxicity to livestock and microbial detoxification. *Journal of Animal Science and Biotechnology*, **8**, 55.

13. Hall, R. J., & Cain, R. B. (1972). Organic fluorine in tropical soils. *New phytologist*, **71**, 839-853.
14. Meyer, J. J., Grobbelaar, N., & Steyn, P. L. (1990). Fluoroacetate-metabolizing pseudomonad isolated from *Dichapetalum cymosum*. *Applied and Environmental Microbiology*, **56**, 2152-2155.
15. Grobbelaar, N., & Meyer, J. M. (1990). Fluoroacetate production by *Dichapetalum cymosum*. *Journal of Plant Physiology*, **135**, 550-553.
16. Meyer, J. M., & O'Hagan, D. (1992). Conversion of fluoropyruvate to fluoroacetate by *Dichapetalum cymosum*. *Phytochemistry*, **31**, 499-501.
17. Meyer, J. M., & O'Hagan, D. (1992). Conversion of 3-fluoropyruvate to fluoroacetate by cell-free extracts of *Dichapetalum cymosum*. *Phytochemistry*, **31**, 2699-2701.
18. Sanada, M., Miyano, T., Iwadare, S., Williamson, J. M., Arison, B. H., Smith, J. L., Douglas, A. W., Liesch, J. M. & Inamine, E. (1986). Biosynthesis of fluorothreonine and fluoroacetate. *The Journal of Antibiotics*, **39**, 259-265.
19. O'Hagan, D., Schaffrath, C., Cobb, S. L., Hamilton, J. T., & Murphy, C. D. (2002). Biochemistry: biosynthesis of an organofluorine molecule. *Nature*, **416**, 279.
20. Schaffrath, C., Deng, H., & O'Hagan, D. (2003). Isolation and characterisation of 5′-fluorodeoxyadenosine synthase, a fluorination enzyme from *Streptomyces cattleya*. *FEBS Letters*, **547**, 111-114.
21. Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X. & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome biology*, **17**, 13.
22. Reymond, P., Weber, H., Damond, M. & Farmer, E. E. (2000). Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. *The Plant Cell*, **12**, 707-19.
23. Mahalingam, R., Gomez-Buitrago, A., Eckardt, N., Shah, N., Guevara-Garcia, A., Day, P., Raina, R. & Fedoroff, N.V. (2003). Characterizing the stress/defense transcriptome of *Arabidopsis*. *Genome Biology*, **4**, 20.
24. Xu, J., Aileni, M., Abbagani, S. & Zhang, P. (2010). A reliable and efficient method for total RNA isolation from various members of spurge family (*Euphorbiaceae*). *Phytochemical Analysis*, **21**, 395-8.
25. Del Fabbro, C., Scalabrin, S., Morgante, M. & Giorgi, F. M. (2013). An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PloS One*, **8**, e85024.

26. Bolger, A.M., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-20.
27. Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M. & MacManes, M. D. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, **8**, 1494.
28. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I. & Pilbout, S. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research*, **31**, 365-70.
29. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T. & Harris, M. A. (2000). Gene ontology: tool for the unification of biology. *Nature Genetics*, **25**, 25.
30. Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., Sunagawa, S., Kuhn, M. & Jensen, L. J. (2015). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*, **44**, 86-93.
31. Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L., Gao, G., Li, C. Y. & Wei, L. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research*, **39**, 16-22.
32. Beier, S., Thiel, T., Münch, T., Scholz, U., & Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics*, **33**, 2583-2585.
33. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A. & Sonnhammer, E. L. (2019). The Pfam protein families database in 2019. *Nucleic acids research*, **47**, 427-432.
34. Bairoch, A. (2000). The ENZYME database in 2000. *Nucleic acids research*, **28**, 304-305.
35. Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., & Gao, G. (2016). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic acids research*, **45**, 1040-1045.
36. Chen, T. W., Gan, R. C., Fang, Y. K., Chien, K. Y., Liao, W. C., Chen, C. C., Wu, T. H., Chang, I. Y., Yang, C., Huang, P. J. & Yeh, Y. M. (2017). FunctionAnnotator, a versatile and efficient web tool for non-model organism annotation. *Scientific reports*, **7**,10430.

37. Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one*, **6**, e21800.
38. Vorwerk, S., Schiff, C., Santamaria, M., Koh, S., Nishimura, M., Vogel, J., Somerville, C. & Somerville, S. (2007). EDR2 negatively regulates salicylic acid-based defenses and cell death during powdery mildew infections of *Arabidopsis thaliana*. *BMC Plant Biology*, 7, 35.
39. Martin, L., Fei, Z., Giovannoni, J., & Rose, J. K. C. (2013). Catalyzing plant science research with RNA-seq. *Frontiers in Plant Science*, **4**, 66.
40. Yang, D., Du, X., Yang, Z., Liang, Z., Guo, Z., & Liu, Y. (2014). Transcriptomics, proteomics, and metabolomics to reveal mechanisms underlying plant secondary metabolism. *Engineering in Life Sciences*, **14**, 456-466.
41. Dong, C., Huang, F., Deng, H., Schaffrath, C., Spencer, J. B., O'Hagan, D. and Naismith, J. H. (2004). Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature*, **427**, 561.
42. Deng, H., Ma, L., Bandaranayaka, N., Qin, Z., Mann, G., Kyeremeh, K., Yu, Y., Shepherd, T. and O'Hagan, D. (2014). Identification of fluorinases from *Streptomyces* sp MA37, *Norcardia brasiliensis*, and *Actinoplanes* sp N902-109 by Genome Mining. *ChemBioChem*, **15**, 364-368.
43. Wang, Y., Deng, Z. and Qu, X. (2014). Characterization of a SAM-dependent fluorinase from a latent biosynthetic pathway for fluoroacetate and 4-fluorothreonine formation in *Nocardia brasiliensis*. *F1000Research*, **3**, 1-13.
44. Ma, L., Li, Y., Meng, L., Deng, H., Li, Y., Zhang, Q. and Diao, A. (2016). Biological fluorination from the sea: discovery of a SAM-dependent nucleophilic fluorinating enzyme from the marine-derived bacterium *Streptomyces xinghaiensis* NRRL B24674. *RSC Advances*, **6**, 27047-27051.
45. Zhu, X., Robinson, D. A., McEwan, A. R., O'Hagan, D., & Naismith, J. H. (2007). Mechanism of Enzymatic Fluorination in *Streptomyces cattleya*. *Journal of the American Chemical Society*, **129**, 14597-14604.
46. Zhao, C., Li, P., Deng, Z., Ou, H. Y., McGlinchey, R. P., & O'Hagan, D. (2012). Insights into fluorometabolite biosynthesis in *Streptomyces cattleya* DSM46488 through genome sequence and knockout mutants. *Bioorganic chemistry*, **44**, 1-7.
47. Murphy, C. D., Moss, S. J., & O'Hagan, D. (2001). Isolation of an Aldehyde Dehydrogenase Involved in the Oxidation of Fluoroacetaldehyde to Fluoroacetate in *Streptomyces cattleya*. *Applied and Environmental. Microbiology*, **67**, 4919-4921.

48. Cai, B., Li, Q., Xu, Y., Yang, L., Bi, H., & Ai, X. (2016). Genome-wide analysis of the fructose 1, 6-bisphosphate aldolase (FBA) gene family and functional characterization of FBA7 in tomato. *Plant physiology and biochemistry*, **108**, 251-265.
49. Brocker, C., Vasiliou, M., Carpenter, S., Carpenter, C., Zhang, Y., Wang, X., Kotchoni, S. O., Wood, A. J., Kirch, H. H., Kopečný, D. & Nebert, D. W. (2013). Aldehyde dehydrogenase (ALDH) superfamily in plants: gene nomenclature and comparative genomics. *Planta*, **237**, 189-210.
50. Su, Z., Łabaj, P. P., Li, S., Thierry-Mieg, J., Thierry-Mieg, D., Shi, W., Wang, C., Schroth, G. P., Setterquist, R. A., Thompson, J. F. & Jones, W. D. (2014) A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nature Biotechnology*, **32**, 903.
51. Zhang, W., Yu, Y., Hertwig, F., Thierry-Mieg, J., Zhang, W., Thierry-Mieg, D., Wang, J., Furlanello, C., Devanarayan, V., Cheng, J. & Deng, Y. (2015). Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biology*, **16**, 133.
52. Kanitz, A., Gypas, F., Gruber, A. J., Gruber, A. R., Martin, G., & Zavolan, M. (2015). Comparative assessment of methods for the computational inference of transcript isoform abundance from RNA-seq data. *Genome Biology*, **16**, 150.
53. Mohanta, T. K., Bashir, T., Hashem, A., Abd_Allah, E. F., Khan, A. L., & Al-Harrasi, A. S. (2018). Early events in plant abiotic stress signaling: Interplay between calcium, reactive oxygen species and phytohormones. *Journal of Plant Growth Regulation*, **37**, 1033-1049.
54. Yan, C., Fan, M., Yang, M., Zhao, J., Zhang, W., Su, Y., Xiao, L., Deng, H. & Xie, D. (2018). Injury activates Ca2+/calmodulin-dependent phosphorylation of JAV1-JAZ8-WRKY51 complex for jasmonate biosynthesis. *Molecular Cell*, **70**, 136-149.
55. Brookes, P. S., Yoon, Y., Robotham, J. L., Anders, M. W., & Sheu, S. S. (2004). Calcium, ATP, and ROS: a mitochondrial love-hate triangle. *American Journal of Physiology-Cell Physiology*, **287**, C817-C833.
56. McCormack, E., Tsai, Y. C., & Braam, J. (2005). Handling calcium signaling: *Arabidopsis* CaMs and CMLs. *Trends in Plant Science*, **10**, 383-389.
57. Das, K., & Roychoudhury, A. (2014). Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. *Frontiers in Environmental Science*, **2**, 53.

58. Snedden, W. A., & Fromm, H. (2001). Calmodulin as a versatile calcium signal transducer in plants. *New Phytologist*, **151**, 35-66.
59. Cheong, Y. H., Kim, K. N., Pandey, G. K., Gupta, R., Grant, J. J., & Luan, S. (2003). CBL1, a calcium sensor that differentially regulates salt, drought, and cold responses in *Arabidopsis. The Plant Cell*, **15**, 1833-1845.
60. Goyal, E., Amit, S. K., Singh, R. S., Mahato, A. K., Chand, S., & Kanika, K. (2016). Transcriptome profiling of the salt-stress response in *Triticum aestivum* cv. Kharchia Local. *Scientific reports*, **6**, 27752.
61. Zhang, S., & Klessig, D. F. (2001). MAPK cascades in plant defense signaling. *Trends in Plant Science*, **6**, 520-527.
62. Bjornson, M., Dandekar, A., & Dehesh, K. (2016). Determinants of timing and amplitude in the plant general stress response. *Journal of Integrative Plant Biology*, **58**, 119-126.
63. Walley, J.W., Coughlan, S., Hudson, M. E., Covington, M. F., Kaspi, R., Banu, G., Harmer, S. L. & Dehesh, K. (2007). Mechanical stress induces biotic and abiotic stress responses via a novel cis-element. *PLoS Genetics*, 3, e172.
64. Wasternack, C., & Hause, B. (2013). Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. *Annals of Botany*, **111**, 1021-1058.
65. Kazan, K. (2015). Diverse roles of jasmonates and ethylene in abiotic stress tolerance. *Trends in Plant Science*, **20**, 219-229.
66. Bari, R., & Jones, J. D. (2009). Role of plant hormones in plant defence responses. *Plant Molecular Biology*, **69**, 473-488.
67. Niki, T., Mitsuhara, I., Seo, S., Ohtsubo, N., & Ohashi, Y. (1998). Antagonistic effect of salicylic acid and jasmonic acid on the expression of pathogenesis-related (PR) protein genes in wounded mature tobacco leaves. *Plant and Cell Physiology*, **39**, 500-507.
68. Lee, A., Cho, K., Jang, S., Rakwal, R., Iwahashi, H., Agrawal, G.K., Shim, J. and Han, O., (2004). Inverse correlation between jasmonic acid and salicylic acid during early wound response in rice. *Biochemical and Biophysical Research Communications*, **318**, 734-738.
69. Sauter, M., Moffatt, B., Saechao, M. C., Hell, R., & Wirtz, M. (2013). Methionine salvage and *S*-adenosylmethionine: essential links between sulfur, ethylene and polyamine biosynthesis. *Biochemical Journal*, **451**, 145-154.

70. Singh, K. B., Foley, R. C., & Oñate-Sánchez, L. (2002). Transcription factors in plant defense and stress responses. *Current Opinion in Plant Biology*, **5**, 430-436.
71. Chen, L., Song, Y., Li, S., Zhang, L., Zou, C., & Yu, D. (2012). The role of WRKY transcription factors in plant abiotic stresses. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, **1819**, 120-128.
72. Vanderauwera, S., Vandenbroucke, K., Inzé, A., Van De Cotte, B., Mühlenbock, P., De Rycke, R., Naouar, N., Van Gaever, T., Van Montagu, M.C. and Van Breusegem, F., (2012). AtWRKY15 perturbation abolishes the mitochondrial stress response that steers osmotic stress tolerance in Arabidopsis. *Proceedings of the National Academy of Sciences*, **109**, 20113-20118.
73. Liu, F., Li, X., Wang, M., Wen, J., Yi, B., Shen, J., Ma, C., Fu, T. and Tu, J., 2018. Interactions of WRKY 15 and WRKY 33 transcription factors and their roles in the resistance of oilseed rape to *Sclerotinia* infection. *Plant Biotechnology Journal*, **16**, 911-925.
74. Chen, H., Lai, Z., Shi, J., Xiao, Y., Chen, Z., & Xu, X. (2010). Roles of *Arabidopsis* WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress. *BMC Plant Biology*, **10**, 281.
75. Zheng, Z., Qamar, S.A., Chen, Z. and Mengiste, T., (2006). *Arabidopsis* WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. *The Plant Journal*, **48**, 592-605.
76. Jiang, Y. and Deyholos, M.K., (2009). Functional characterization of *Arabidopsis* NaCl-inducible WRKY25 and WRKY33 transcription factors in abiotic stresses. *Plant Molecular Biology*, **69**, 91-105.
77. Hu, Y., Dong, Q., & Yu, D. (2012). Arabidopsis WRKY46 coordinates with WRKY70 and WRKY53 in basal resistance against pathogen *Pseudomonas syringae*. *Plant Science*, **185,** 288-297.
78. Ding, Z. J., Yan, J. Y., Xu, X. Y., Yu, D. Q., Li, G. X., Zhang, S. Q., & Zheng, S. J. (2014). Transcription factor WRKY 46 regulates osmotic stress responses and stomatal movement independently in *Arabidopsis*. *The Plant Journal*, **79**, 13-27.
79. Pandey, S. P., Srivastava, S., Goel, R., Lakhwani, D., Singh, P., Asif, M. H., & Sane, A. P. (2017). Simulated herbivory in chickpea causes rapid changes in defense pathways and hormonal transcription networks of JA/ethylene/GA/auxin within minutes of wounding. *Scientific Reports*, **7**, 44729.

80. Pateraki, I. and Kanellis, A.K., (2010). Stress and developmental responses of terpenoid biosynthetic genes in *Cistus creticus* subsp. creticus. *Plant cell reports*, **29**, 629-641.
81. Bede, J. C., Musser, R. O., Felton, G. W., & Korth, K. L. (2006). Caterpillar herbivory and salivary enzymes decrease transcript levels of *Medicago truncatula* genes encoding early enzymes in terpenoid biosynthesis. *Plant Molecular Biology*, **60**, 519-531.
82. Liu, R., Xu, S., Li, J., Hu, Y. and Lin, Z., 2006. Expression profile of a PAL gene from *Astragalus membranaceus* var. Mongholicus and its crucial role in flux into flavonoid biosynthesis. *Plant Cell Reports*, **25**, 705-710.

Functional characterisation of the transcriptome from the fluoroacetate-producing plant, *Dichapetalum cymosum*, in response to mechanical wounding

Selisha A. Sooklal[1,+] , Phelelani Mpangase[2,+] , Shaun Aaron[2] , Robert H. Archer[3] and Karl Rumbold[1,*]

[1]Department of Microbiology, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, 2000, South Africa

[2]Sydney Brenner Institute for Molecular Biosciences, University of the Witwatersrand, Johannesburg, 2000, South Africa

[3]National Herbarium, South African National Biodiversity Institute, Pretoria, 0186, South Africa

[*]Corresponding author e-mail: karl.rumbold@wits.ac.za

[+]These authors contributed equally to this work

Supplementary Information

Table S1: Quantification of total RNA extracted from *D. cymosum* leaves before and after mechanical wounding. RNA was quantified using the Qubit Fluorometer 2.0.

	T:0 (control) unwounded	**T:30 post-wounding**	**T:60 post-wounding**
Replicate 1	368 ng/µl	97 ng/µl	263 ng/µl
Replicate 2	243 ng/µl	41 ng/µl	176 ng/µl
Replicate 3	149 ng/µl	28 ng/µl	171 ng/µl

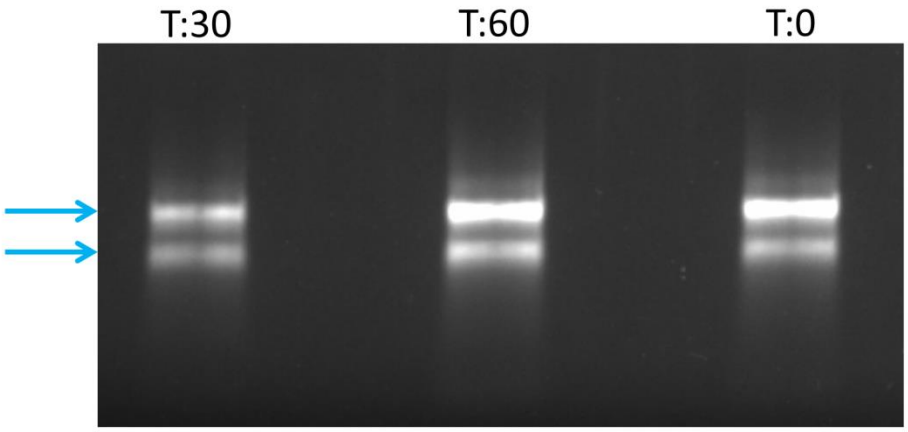
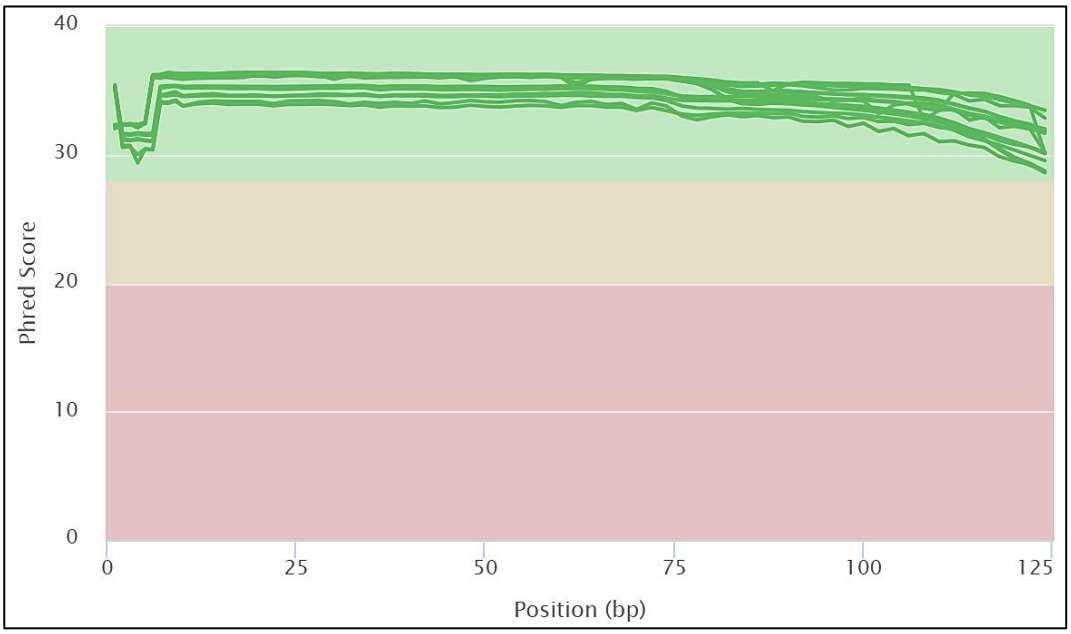


Figure S1: Agarose (1%) gel electrophoresis showing the integrity of the total RNA extracted from *D. cymosum* leaves.

Table S2. Table providing the link and corresponding reference for the tools and databases used in this study

Tool or database	Link	Reference (if available)
FastQC	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/	-
Trimmomatic	http://www.usadellab.org/cms/?page=trimmomatic	[1]
Trinity	https://github.com/trinityrnaseq/trinityrnaseq/wiki	[2]
SwissProt	www.uniprot.org	[3]
TrEMBL	www.uniprot.org	[3]
GO	http://geneontology.org/	[4]
eggNOG	http://eggnogdb.embl.de/#/app/home	[5]
Pfam	https://pfam.xfam.org/	[6]
ExPASy-Enzyme	https://enzyme.expasy.org/	[7]
MISA	https://webblast.ipk-gatersleben.de/misa/	[8]
PlantTFDB	http://planttfdb.cbi.pku.edu.cn/	[9]
KOBAS	http://kobas.cbi.pku.edu.cn/	[10]
Function Annotator	http://fa.cgu.edu.tw/	[11]
REVIGO	http://revigo.irb.hr/	[12]

(A)



(B)

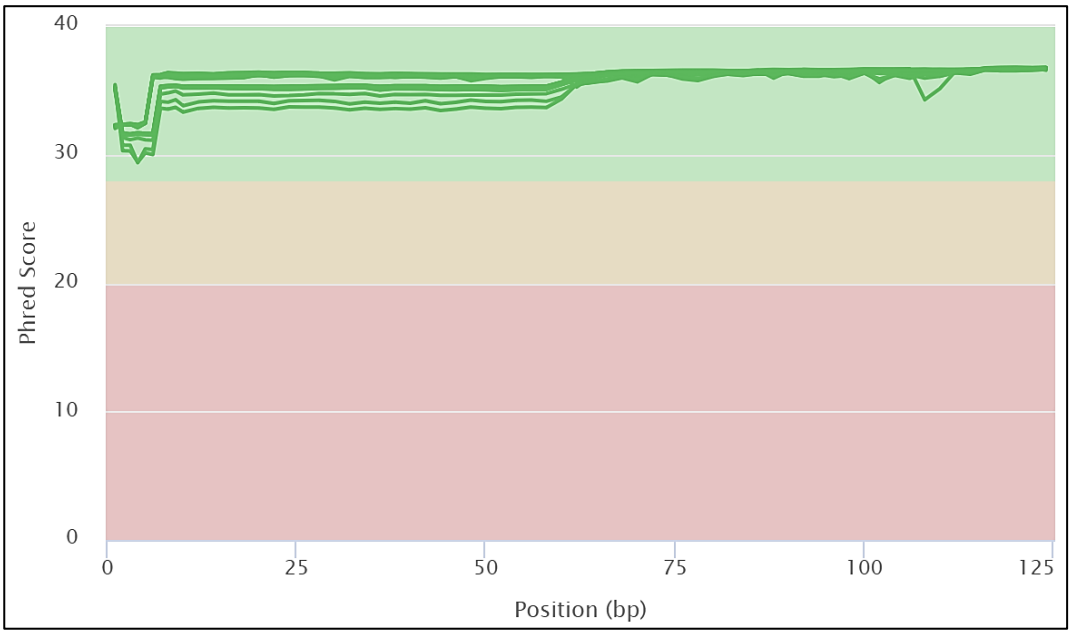


Figure S2. MultiQC report showing the Phred score for each sequenced library of the unwounded and wounded *D. cymosum* leaves (A) before trimming and (B) after trimming using Trimmomatic.

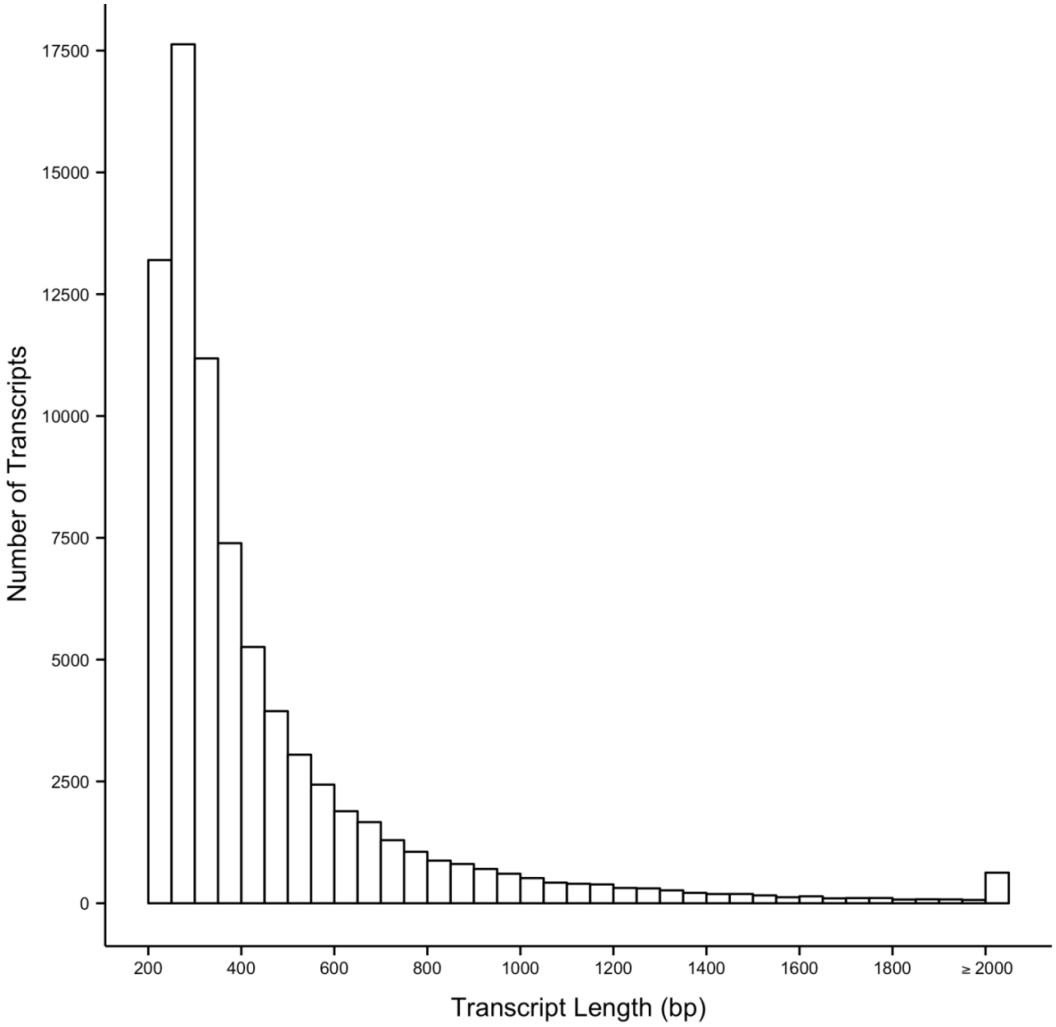


Figure S3. Length distribution of assembled transcripts. The 77,845 transcripts assembled with Trinity had an average length of 454 bp. Majority of transcripts fell between 200-500 bp while the shortest transcript had a length of 224 bp and the longest transcript had a length of 10,795 bp.

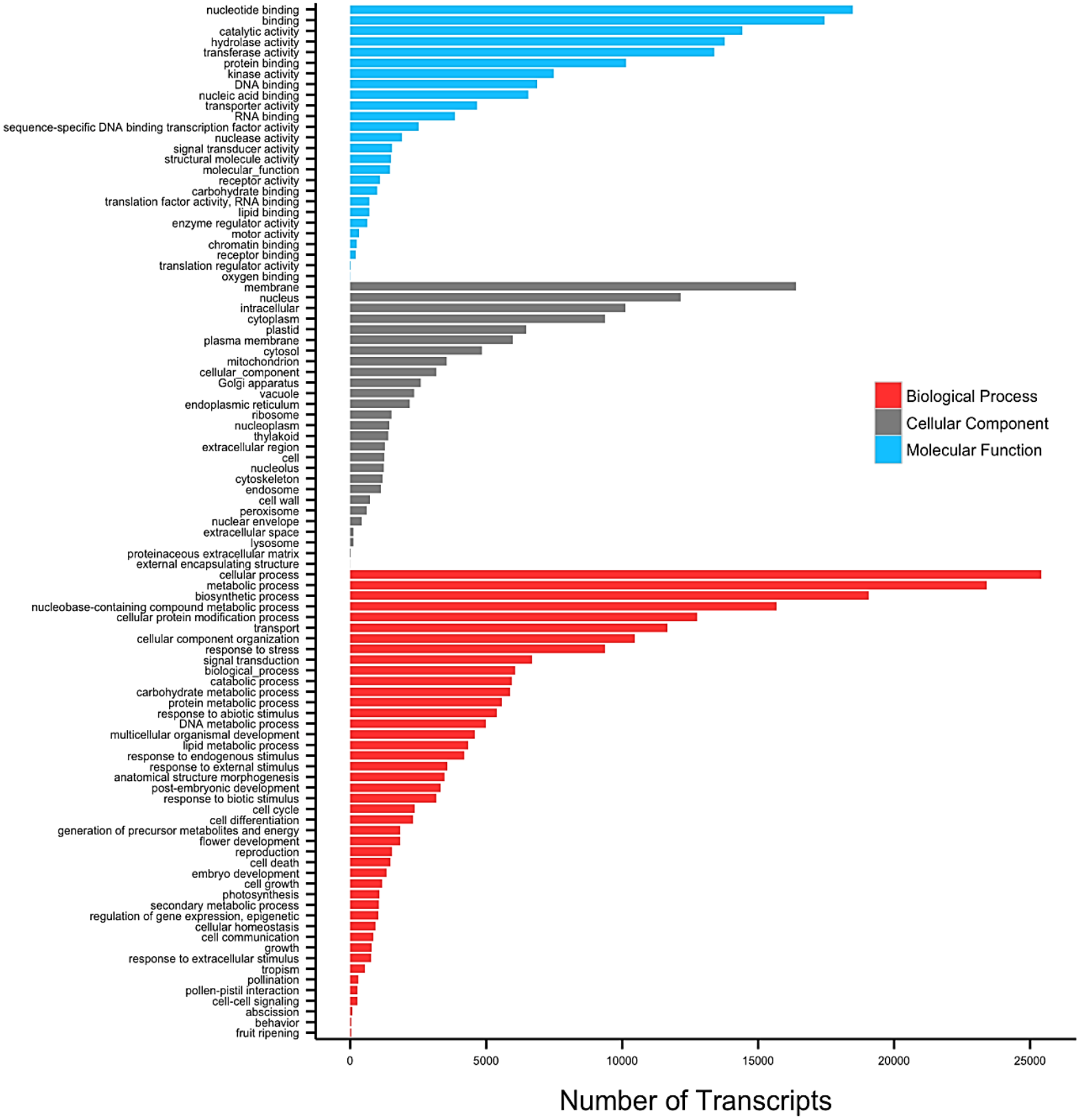


Figure S4. GO classification of the *D. cymosum* transcriptome. GO classified transcripts into three broad domains: biological process (red), cellular component (grey) and molecular function (blue). Annotations were further sub-divided using 97 plant specific GO Slims (listed on the y-axis).

Table S3. Detection of SSR's in the *D. cymosum* transcriptome. SSR's were found using the MISA tool

	Number of SSR's	Percentage of total SSR's	Most frequent repeat
di-	393	25.6	AG/CT
tri-	887	57.8	CTT/AAG
tetra-	179	11.7	AAAT/ATTT
penta-	50	3.3	AGTTT/AAACT
hexa-	25	1.6	TGTGGC/GCCACA

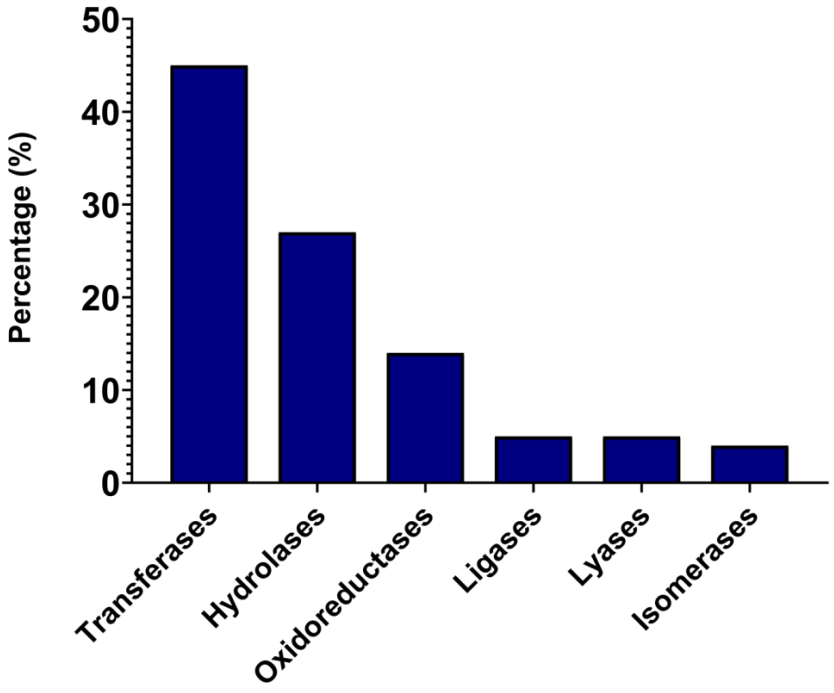


Figure S5. EC classification of 7,311 *D. cymosum* transcripts into the oxidoreductase (14%), transferase (45%), hydrolase (27%), lyase (5%), isomerase (4%) and ligase (5%) categories.

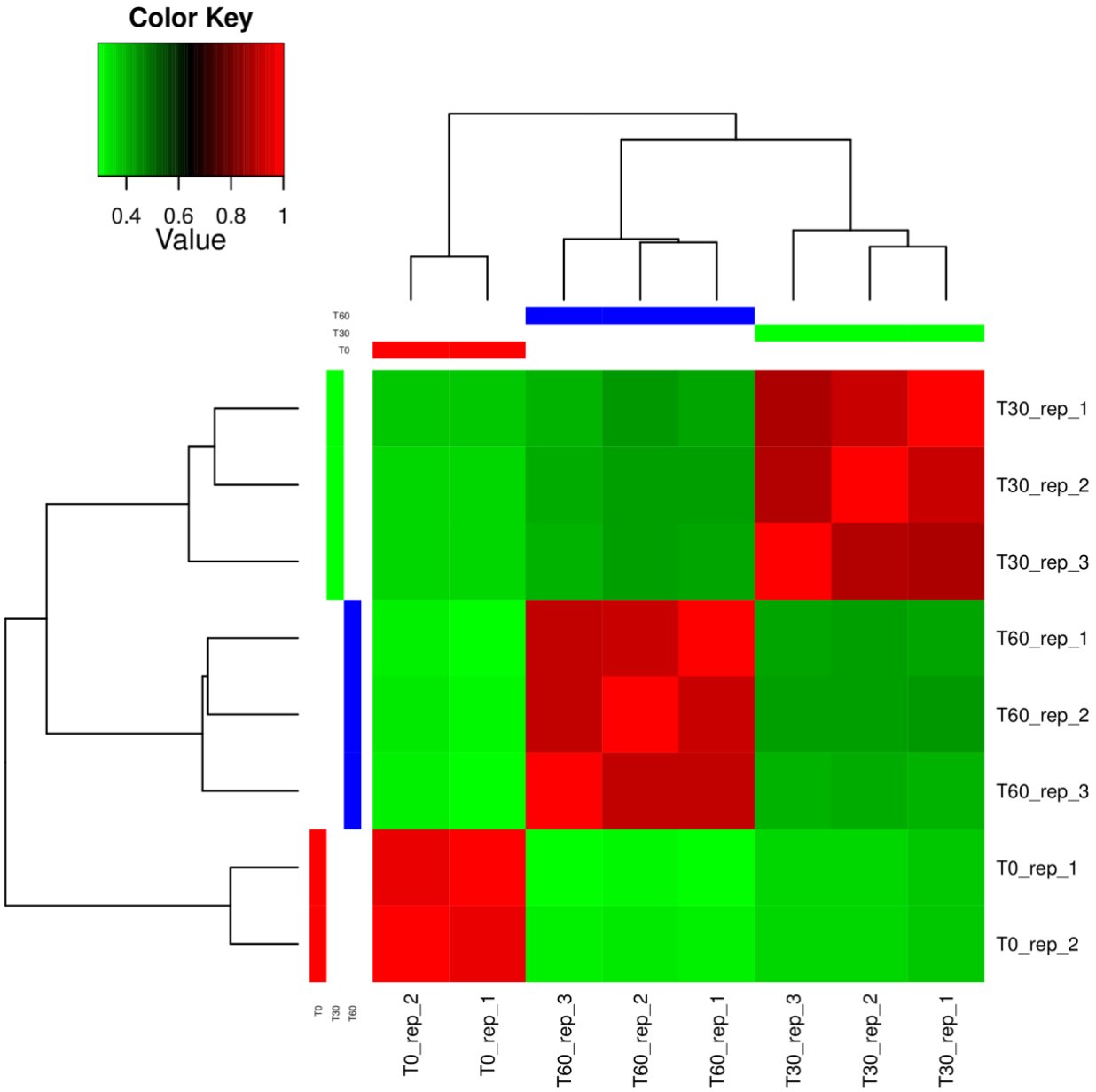


Figure S6. Correlation matrix between replicates of each time point. A perfect correlation (1) is indicated in red while lower correlation (0.4) is indicated in green. Replicates within each time point shows good correlation indicating reproducibility of experimental data.

Table S4. Blastp results following the query of the *S. cattleya* fluorometabolite biosynthetic genes against the *D. cymosum* transcriptome

Query ID	Hit in transcriptome (Subject ID)	% of identical matches	Alignment length	No. of mis-matches	No. of gap openings	Start-end of alignment in query	Start-end of alignment in subject	e-value	Bit score	Putative function of *D. cymosum* transcript % identity, query coverage, e-value
WP_014144878.1 **Fluorinase**	TR6001_c0_g2_i4_(-2)	36.8	38	23	1	155-192	36-72	5.6	29.3	**Ceramide inositol-phosphotransferase** 93.0%, 73%, 7e-40
WP_014144879.1 **PNP**	TR35484_c0_g1_i1_(-2)	29.6	54	35	1	167-220	1-51	1.4	31.2	**LRR receptor-like serine/threonine-protein kinase** 76.6%, 100%, 1e-50
WP_014142773.1 **Isomerase**	TR31930_c0_g1_i1_(-2)	26.5	253	174	6	84-333	172-415	3.91E-12	69.3	**IF-2B domain-containing protein** 90.9%, 77%, 0.0
WP_014141855.1 **Aldolase**	TR16309_c0_g2_i1_(-1)	55.9	256	110	2	112-367	2-254	3.17E-81	252	**Uncharacterised protein: transaldolase family** 92.6%, 85%, 9e-169
WP_014150905.1 **Aldolase**	TR25733_c0_g1_i1_(-2)	38.4	99	59	1	45-143	48-144	5.22E-15	73.6	**Uncharacterised protein: putative dehydrogenase** 79.3%, 72%, 2e-62
WP_014141713.1 **Aldehyde dehydrogenase**	TR21022_c0_g1_i3_(-1)	41.5	482	259	10	24-496	114-581	1.64E-112	348	**Aldehyde dehydrogenase family 2 member** 85.2%, 86%, 0.0
WP_014151017.1 **4-FT transaldolase**	TR44173_c0_g2_i1_(-1)	21.7	423	270	11	17-395	41-446	1.93E-13	75.1	**Serine hydroxymethyltransferase** 87.3%, 100%, 0.0

Table S5. The regulation of transcripts in response to mechanical wounding in *D. cymosum*

Category	Putative fuction	Regulation			Transcript ID
		T:0	T:30	T:60	
Transcription factors					
B3	Auxin response factor 11	↑	↓	↑	TR9948_c0_g1_i2
bHLH	Transcription factor bHLH13	↓	↑	↑	TR43985_c0_g2_i1
	Transcription factor bHLH148	↑	•	↓	TR44545_c0_g1_i1
C2H2	Zinc finger protein ZAT10	↓	↑	•	TR31532_c0_g1_i1
C3H	Zinc finger CCCH domain−containing protein 29	↓	↑	↑	TR1170_c0_g1_i1 TR1170_c0_g1_i2 TR6469_c1_g1_i1
	Zinc finger CCCH domain−containing protein 47	↓	↑	↑	TR9322_c0_g1_i1
	Zinc finger CCCH domain−containing protein 33	↓	↑	↑	TR861_c0_g2_i3
CAMTA	Calmodulin−binding transcription activator 2	↓	•	↑	TR2498_c0_g2_i4
Co-like	Zinc finger protein CONSTANS−LIKE 16	↑	↑	↓	TR25717_c0_g1_i1 TR25717_c0_g2_i1
Dof	Cyclic dof factor 3	↑	↑	↓	TR25324_c0_g1_i1
ERF	Ethylene−responsive transcription factor RAP2−4	↓	↑	↑	TR10154_c3_g2_i1
	Ethylene−responsive transcription factor ERF109	↓	↑	↑	TR10345_c0_g1_i1
	Ethylene−responsive transcription factor ERF105	↓	↑	↑	TR28723_c1_g1_i1
GRAS	Chitin−inducible gibberellin−responsive protein 1	↓	↑	•	TR31813_c1_g1_i1
	Scarecrow−like protein 14	↓	↑	↓	TR42270_c0_g1_i1 TR680_c0_g1_i1
MIKC_MADS	MADS−box protein SVP	↑	↑	↓	TR14910_c0_g1_i3
M-type_MADS	MADS−box protein JOINTLESS	↑	↑	↓	TR8542_c0_g4_i7
MYB	Myb−related protein 306	↑	↑	↓	TR34592_c0_g1_i1
MYB_related	Protein REVEILLE 1	↑	•	↓	TR39440_c1_g1_i2
NAC	NAC transcription factor ONAC010	↓	↑	•	TR11949_c0_g1_i2
	NAC domain−containing protein 78	↓	↑	↑	TR24150_c0_g1_i3
	NAC domain−containing protein 86	↓	↑	•	TR6271_c0_g2_i4
NF-YC	Nuclear transcription factor Y subunit C−2	↓	↓	↑	TR30184_c0_g1_i1
Trihelix	Trihelix transcription factor ASIL2	↑	•	↓	TR45736_c0_g1_i1
WRKY	Probable WRKY transcription factor 53	↓	↑	•	TR1133_c1_g1_i1 TR1133_c1_g1_i2 TR1133_c1_g1_i3
	WRKY transcription factor 6	↓	↑	↑	TR21652_c0_g1_i3
	Probable WRKY transcription factor 33	↓	↑	•	TR27647_c4_g1_i3
	Probable WRKY transcription factor 46	↓	↑	↑	TR31296_c0_g1_i4 TR31296_c0_g1_i5 TR34179_c0_g1_i3
	Probable WRKY transcription factor 15	↓	↑	↑	TR41435_c3_g1_i2
	WRKY transcription factor 18	↓	↑	↑	TR43264_c1_g1_i1
Phytohormone responses					
Jasmonates	Linoleate 13S−lipoxygenase	↓	↑	↑	TR33072_c0_g2_i1
	12−oxophytodienoate reductase 3	↓	↑	↑	TR37738_c0_g1_i5 TR37738_c0_g1_i2
	Protein TIFY 6B (JAZ3)	↓	↓	↑	TR38457_c1_g1_i1
	Protein TIFY 10A (JAZ1)	↓	↑	↑	TR31372_c1_g2_i2 TR24873_c0_g1_i1 TR24873_c0_g1_i3
	Protein TIFY 11B (JAZ6)	↓	↑	↑	TR30411_c0_g2_i1
Ethylene	*S*-adenosylmethionine synthase 2	↓	↑	↑	TR6199_c0_g1_i1 TR6199_c0_g1_i2
	1−aminocyclopropane−1−carboxylate oxidase	↓	↑	↑	TR22364_c0_g1_i2

Gibberellins	Gibberellin 2−beta−dioxygenase 2	↓	↑	↓	TR3986_c1_g1_i1
	Gibberellin receptor GID1	↑	↑	↓	TR5384_c0_g1_i1
	DELLA protein GAI	↑	↓	•	TR37799_c0_g1_i1
Auxins	IAA−amino acid hydrolase ILR1−like 4	↓	↑	↑	TR44491_c0_g1_i1
	Auxin−repressed 12.5 kDa protein	↑	↑	↓	TR29771_c0_g3_i1
	Auxin-responsive protein IAA16	↑	•	↓	TR17243_c0_g1_i1
	Tryptophan aminotransferase-related protein 1	↑	↓	↓	TR12889_c0_g1_i1 TR12889_c0_g1_i2
	Tryptophan aminotransferase-related protein 1	↑	↑	↓	TR38367_c11_g13_i2 TR38367_c11_g20_i1 TR8487_c0_g2_i4 TR23600_c0_g2_i1
Salicylic acid	Protein ENHANCED DISEASE RESISTANCE 2	↑	↑	↓	TR43024_c0_g1_i10
Abscisic acid	Abscisic stress−ripening protein 3	↑	•	↓	TR21333_c0_g1_i1
Heat shock proteins					
	Heat shock 70 kDa protein 2	↓	•	↑	TR39531_c1_g1_i1
	Heat shock 70 kDa protein 4	↓	•	↑	TR24193_c0_g1_i1
MAP kinases					
	Mitogen−activated protein kinase kinase kinase MPK1	↓	↑	↑	TR32256_c0_g1_i1
	Mitogen-activated protein kinase 3	↓	↑	•	TR41832_c1_g1_i1 TR41832_c1_g1_i2 TR41832_c1_g1_i3
Secondary metabolite biosynthesis					
Terpenoid biosynthesis	1−deoxyxylulose−5−phosphate synthase	↓	↓	↑	TR36209_c0_g2_i1
	Isoprene synthase	↓	↓	↑	TR25944_c2_g1_i2 TR25944_c2_g1_i8
Ca^{2+} signalling					
	Calcium−transporting ATPase 1	↓	↑	↑	TR12376_c2_g6_i2
	Calcium−transporting ATPase 2	↓	↑	↑	TR12376_c1_g1_i1 TR12376_c1_g1_i2
	Calcium−transporting ATPase 12	↓	↑	↑	TR5940_c0_g1_i1
	Calmodulin-like protein CML45	↓	↑	•	TR37317_c0_g1_i1
	Calmodulin-like protein CML27	↓	↑	↑	TR24640_c0_g1_i2
	Calmodulin (CaM)	↓	↑	↑	TR4728_c0_g1_i2
	Calmodulin−1 (CaM−1)	↓	↑	•	TR36505_c0_g1_i2
	Calmodulin−binding transcription activator 4	↓	•	↑	TR25454_c0_g2_i1
	Calcineurin B−like protein 4	↑	•	↓	TR18111_c0_g2_i1
	Multiple C2 and transmembrane domain−containing protein 1	↓	↑	↑	TR13280_c0_g1_i1
Transporters					
	Boron transporter 4	↓	↓	↑	TR5388_c1_g1_i8
	Copper transport protein ATX1	↑	↓	↓	TR36133_c0_g1_i2 TR36133_c0_g1_i1
	Probable aquaporin PIP1−2	↓	↑	↑	TR27512_c4_g1_i1
	Probable inorganic phosphate transporter 1−5	↓	↑	↑	TR41082_c0_g1_i2
	Vacuolar amino acid transporter 1	↓	↑	↑	TR13739_c0_g1_i3 TR13739_c0_g1_i4
	Protein TRANSPARENT TESTA 12	↓	↑	↑	TR22329_c0_g1_i2
Proteolysis					
	E3 ubiquitin−protein ligase PUB23	↑	↓	↓	TR16704_c0_g1_i1
	E3 ubiquitin−protein ligase RING1−like	↓	↑	↓	TR42553_c0_g1_i1
	Probable E3 ubiquitin−protein ligase HERC1	↑	↑	↓	TR16704_c0_g1_i1
	Polyubiquitin 11	↓	•	↑	TR30506_c1_g1_i5

					TR30506_c1_g1_i6 TR7579_c0_g3_i2
	Ubiquitin carboxyl−terminal hydrolase 12	↓	↑	↓	TR12222_c1_g1_i2
	Putative U−box domain−containing protein 42	↓	↑	↑	TR33493_c0_g1_i1
	Aspartic proteinase A1	↑	•	↓	TR39950_c0_g2_i1
	Aspartic proteinase−like protein 1	↓	↑	↑	TR21898_c0_g1_i3
	ATP−dependent zinc metalloprotease	↓	↑	↓	TR7959_c0_g2_i1
	Basic secretory protease	↓	↑	↑	TR7316_c1_g1_i1
	Thiol protease aleurain	↑	↑	↓	TR39522_c0_g3_i1
	Rhomboid protease GluP	↑	↓	↓	TR35072_c0_g1_i1
Mitochondrial respiratory chain					
	NADH dehydrogenase subunit 1	↓	↑	↓	TR23950_c0_g1_i1
	NADH dehydrogenase subunit 2	↓	↑	↓	TR25849_c0_g1_i1 TR6136_c0_g1_i1 TR12975_c2_g5_i5 TR12975_c2_g9_i1
	NADH dehydrogenase subunit 3	↓	↑	↓	TR4733_c0_g1_i1
	NADH dehydrogenase subunit 4	↓	↑	↓	TR13996_c0_g1_i1
	NADH dehydrogenase subunit 5	↓	↑	↓	TR39326_c0_g1_i1
	NADH dehydrogenase subunit 6	↓	↑	↓	TR10815_c0_g1_i1
	NADH dehydrogenase subunit 7	↑	↓	•	TR742_c2_g1_i1 TR13536_c2_g2_i4
	Cytochrome c oxidase subunit 1	↓	↑	↓	TR13956_c0_g1_i1
	Cytochrome c oxidase subunit 2	↓	↑	↓	TR6576_c0_g1_i1
	Cytochrome c oxidase subunit 3	↓	↑	↓	TR44521_c0_g1_i1
	Cytochrome b	↓	↑	↓	TR45100_c0_g1_i1
Photorespiration					
	Cytochrome b6	↓	↑	↑	TR33038_c0_g1_i1 TR10321_c1_g1_i1
	Cytochrome b6-f complex subunit 8	↓	↑	↑	TR30276_c0_g1_i1
	Ferredoxin-2	↓	↑	↑	TR38279_c0_g1_i1
	Photosystem II protein D1	↓	↑	↑	TR45648_c2_g1_i1 TR45648_c0_g2_i1
	Photosystem II D2	↓	↓	↑	TR33459_c0_g3_i2
	Photosystem II CP47 reaction centre protein	↓	↑	↓	TR2524_c1_g1_i1
	Photosystem II reaction centre protein H	↓	↑	↑	TR10321_c0_g1_i1
	ATP synthase	↓	↑	↑	TR38824_c0_g1_i2 TR26918_c2_g5_i1
	Photosystem I assembly protein Ycf3	↓	↑	↑	TR9274_c0_g2_i2
	Tetrapyrrole−binding protein	↑	↓	↓	TR10917_c0_g1_i1
Oxidative stress					
	Catalase isozyme 2	↓	↑	↑	TR19830_c0_g1_i1
	Superoxide dismutase	↑	↓	↓	TR39709_c0_g1_i1
	Glutaredoxin−C9	↓	↑	•	TR5452_c0_g1_i1
	Reticuline oxidase−like protein	↓	↑	↑	TR37436_c0_g1_i1
	Thioredoxin F−type	↓	↓	↑	TR43948_c1_g1_i4
	NAD(P)H-quinone oxidoreductase	↓	↑	↑	TR37887_c0_g1_i1
Phosphatases					
	Purple acid phosphatase 3	↓	↑	↑	TR40803_c0_g2_i6
	Probable protein phosphatase 2C 24	↑	•	↓	TR44379_c0_g1_i2
	Probable protein phosphatase 2C 25	↓	↑	•	TR31269_c0_g1_i1
	Protein phosphatase 2C 37	↑	↑	↓	TR26183_c0_g1_i1
	Probable protein phosphatase 2C 44	↓	↑	↓	TR5165_c0_g2_i2
	Probable protein phosphatase 2C 63	↓	↑	↑	TR33687_c0_g1_i9

Pathogenesis-related					
	Protein argonaute 5	↓	•	↑	TR19756_c0_g1_i1
	Probable disease resistance protein	↓	↑	•	TR28205_c1_g1_i1 TR28205_c1_g1_i2
	Indole−3−acetic acid−induced protein ARG2	↓	↑	•	TR20755_c0_g1_i1
	MLP−like protein 31	↓	↑	↑	TR39328_c0_g1_i1
	Pleiotropic drug resistance protein 2	↓	↑	↓	TR42006_c0_g2_i4 TR34682_c0_g1_i3
	Pleiotropic drug resistance protein 2	↓	↑	↑	TR42006_c0_g2_i2 TR34682_c0_g1_i2
	Putative late blight resistance protein homolog R1B−19	↓	↑	↓	TR22540_c0_g1_i2
Cell wall – signalling and modification					
	Remorin (pp34)	↑	↑	↓	TR10463_c1_g1_i2
	Protein EXORDIUM	↓	↑	•	TR29005_c0_g1_i1
Other kinases					
	Probable receptor−like protein kinase	↓	↑	↓	TR7041_c0_g1_i1 TR7041_c0_g1_i2
	Serine/threonine−protein kinase prp4	↑	•	↓	TR7054_c0_g1_i1
	Leucine−rich repeat receptor−like serine/threonine/tyrosine−protein kinase	↓	↑	↑	TR44916_c0_g1_i1
	Glycerophosphodiester phosphodiesterase protein kinase domain−containing GDPDL2	↓	↑	↓	TR46484_c0_g1_i2
	Putative 1−phosphatidylinositol−3−phosphate 5−kinase FAB1D	↑	↑	↓	TR34194_c0_g1_i2
Genetic information processing					
	CCR4-associated factor 1 homolog	↓	↑	•	TR43774_c0_g1_i1
	DNA helicase	↑	↑	↓	TR3254_c0_g1_i1
	Putative transcription elongation factor SPT5	↑	↓	↓	TR33461_c0_g1_i1
	Protein translation factor SUI1 homolog	↓	↑	↑	TR2718_c0_g1_i1
	Serine/arginine−rich splicing factor RS40	↑	↓	↑	TR3607_c0_g1_i2
	Eukaryotic translation initiation factor 3 subunit A	↑	↓	↓	TR22446_c0_g1_i3
	CLK4−associating serine/arginine rich protein	↑	↑	↓	TR31855_c0_g2_i1
	Far upstream element−binding protein 3	↑	↓	↓	TR1721_c0_g1_i1
	Regulator of rDNA transcription protein 15	↓	↑	↓	TR44904_c0_g2_i1 TR8384_c0_g1_i1
	Heterogeneous nuclear ribonucleoprotein 1	↑	•	↓	TR36557_c0_g1_i1
	Histone H3.3	↑	↑	↓	TR28838_c2_g1_i1
	FIP1[V]−like protein	↑	↑	↓	TR26809_c1_g2_i2

↑ **Up-regulated,** ↓ **Down-regulated,** • **Fold-change ≥ -1 and ≤ 1**

References

1. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.
2. Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M. & MacManes, M. D. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8, 1494.
3. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I. & Pilbout, S. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research*. 31, 365-70.
4. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T. & Harris, M. A. (2000). Gene ontology: tool for the unification of biology. *Nature Genetics*, **25**, 25.
5. Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., Sunagawa, S., Kuhn, M. & Jensen, L. J. (2015). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*, **44**, 86-93.
6. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A. & Sonnhammer, E. L. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, **47**, 427-432.
7. Bairoch, A. (2000). The ENZYME database in 2000. *Nucleic Acids Research*, **28**, 304-305.
8. Beier, S., Thiel, T., Münch, T., Scholz, U., & Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics*, **33**, 2583-2585.
9. Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., & Gao, G. (2016). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, **45**, 1040-1045.
10. Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L., Gao, G., Li, C. Y. & Wei, L. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research*, **39**, 16-22.
11. Chen, T. W., Gan, R. C., Fang, Y. K., Chien, K. Y., Liao, W. C., Chen, C. C., Wu, T. H., Chang, I. Y., Yang, C., Huang, P. J. & Yeh, Y. M. (2017). FunctionAnnotator, a

versatile and efficient web tool for non-model organism annotation. *Scientific Reports*, **7**,10430.

12. Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS One*, **6**, e21800.

Chapter 3

Proteomic profiling and integration of transcriptomic data for *Dichapetalum cymosum*

Introduction

Dichapetalum cymosum is one of the few plants worldwide that has been found to produce fluorometabolites [1,2]. In 1944, fluoroacetate (FA) was isolated as the toxic component of *D. cymosum* [1]. Although all parts of the plant contain FA, the highest concentrations were observed in the young leaves, seeds and flowers of *D. cymosum*. The young leaves accumulate FA at concentrations up to 232 $mg \cdot kg^{-1}$ fresh mass, while the seeds and flowers contain up to 410 $mg \cdot kg^{-1}$ and 362 $mg \cdot kg^{-1}$, respectively [3]. The biosynthesis of FA by *D. cymosum* led to conjecture of a fluorinating enzyme [4]. However, such an enzyme remains elusive and the biosynthetic origin of the C-F bond has not been elucidated in higher plants.

Since the discovery of the first fluorinating enzyme (or fluorinase) from *S. cattleya*, new information has become available such as the gene sequence and natural substrate of bacterial fluorinases [5,6]. In the previous chapter (Chapter 2), the transcriptome of *D. cymosum* was sequenced and investigated for transcripts with similarity to bacterial fluorinases. However, a transcript with homology was not found. Zhu et al had proposed the evolution of a different biological mechanism for fluoride incorporation in plants [7]. The findings from the *D. cymosum* transcriptome may support this hypothesis. However, if this is true, then discovery approaches based on sequence similarity will inherently fail. Moreover, it could also imply that that a new sequence may be present in the transcriptome but not identified.

Shotgun proteomics has become the preferred method for detection and profiling of proteins [8]. It allows for the analysis of complex protein samples with high efficiency and avoids limitations, such as bias against some proteins and reproducibility issues, associated with 2-dimensional gel electrophoresis (2DE) coupled to mass spectrometry (MS) [9]. In shotgun

proteomics, proteins samples are enzymatically digested. Peptides are separated using liquid chromatography and then analysed using mass spectrometry [10]. Proteins are inferred by comparing the mass-to-charge ratio (*m/z*) of peptides to public databases. However, identification is still limited to only known proteins and this is problematic for the identification of novel proteins. In such cases, a custom database comprising of information, such as the genome or transcriptome, from the species of interest is necessary [10].

Integration of the omics can provide vital information that is usually otherwise missed during separate individual analysis. Lanfranchi et al reported the identification of a new class of HNL from fern plants through the integration of omic data coupled with functional screening [11]. The HNL enzyme was undetectable in the transcriptome. However, when protein fractions were assayed, activity was detected. These protein fractions were characterized using mass-spectrometry (MS) and a proteome was generated. When searched against the transcriptome, 36 protein sequences were identified. The list was narrowed down until 6 candidates were synthesized into expression plasmids in *E. coli*. This led to the identification of a new class of HNL enzyme with unique folding patterns [11].
The synergism between omic integration and functional screening can allow for the identification of truly novel enzymes [10]. Therefore, in this study, the proteome of *D. cymosum* leaf, seed and flower tissues were generated using LC-MS/MS and integrated with transcriptomic data to investigate the presence of a potential fluorinating enzyme.

Materials and methods

Materials

The *S*-adenosyl-$_{L}$-methionine (SAM), NaF and HPLC-grade solvents were purchased from Sigma-Aldrich (Missouri, USA). The cOmplete protease inhibitor cocktail was from Roche (Basel, Switzerland) and the modified trypsin was from Promega (Madison, USA). The P-PER Plant Protein Extraction Kit and Zebra Spin desalting columns were from Thermo Fisher Scientific (Massachusetts, USA). The 5 kDa MWCO Spin-X UF 500 centrifugal concentrators purchased from Corning (New York, USA). The Bio-Rad mini-PROTEAN gel electrophoresis system was used for SDS-PAGE (California, USA). The synthetic 5'-fluoro-5'-deoxyadenosine (5'-FDA) sample was a kind gift from Dr. Allan Prior (now at the University of Hawaii). All other reagents were purchased from either VWR (Pennsylvania, USA) or SRL (Mumbai, India).

Collection of plant material

The flower, fruit and young leaf tissue of *Dichapetalum cymosum* (gifblaar) were collected at the Pretoria National Botanical Gardens, South Africa, in October 2017 during the spring season. Samples were immediately frozen in liquid nitrogen and later transferred to a - 80 °C freezer until further processing.

Protein extraction and fractionation

Total protein from the young leaf (~ 100 mg) was isolated using the P-PER Plant Protein Extraction Kit. The seed was obtained from inside the fruit following removal of the pericarp. However, insufficient protein yields were obtained for both the seed and flower using the P-PER Protein Extraction Kit. Therefore, tissue was ground to a fine powder in liquid nitrogen using a pre-chilled mortar and pestle. The powder obtained from the seed (~ 1 g) and flower (~ 200 mg) were resuspended in 5 volumes (w/v) of extraction buffer (50 mM sodium borate, pH 8.5, 2 mM EDTA, 5% PVPP, 5 mM DTT, cOmplete protease inhibitor cocktail) and incubated for 20 min at 4 °C with occasional vortexing. Plant debris was removed by centrifuging at 13,000 x *g* for 15 min at 4 °C. Prior to denaturation using phenol, activity was assayed using the total protein obtained from the young leaf, seed and flower (as described in 'fluorinase activity assay'). A phenol-based method adapted from Vincent et al was then used to extract protein from the supernatant [12]. In brief, an equal volume of tris-buffered phenol (pH 7.9) was added to the supernatant and incubated at - 20 °C for 30 min with occasional vortexing. The solution was centrifuged at 10,000 x *g* for 15 min at 4 °C and the phenol phase was transferred to a new tube. This was extracted once more with an equal volume of buffer (50 mM sodium borate, pH 8.5, 2 mM EDTA, 5 mM DTT, cOmplete protease inhibitor cocktail) and centrifuged as before. The phenol phase was collected to which 5 volumes of pre-chilled 0.1 M ammonium acetate in methanol was added and proteins were precipitated overnight at - 20 °C. Precipitated proteins were collected by centrifugation at 12,000 x *g* for 10 min at 4 °C. The resultant pellet was washed twice with cold 0.1 M ammonium acetate in methanol and once with cold 10 mM DTT in acetone. The protein pellets for the young leaf, seed and flower were air-dried and resuspended in SDS-PAGE sample buffer (150 mM tris-HCl, pH 7.0, 12%

SDS, 6% β-mercaptoethanol, 30% glycerol, 0.05% Coomassie blue-G250) as described by Schägger [13].

Fractionation of total protein, based on charge, was also performed to enrich low abundant protein species. *D. cymosum* leaf tissue (~ 2.7 g) was crushed in liquid nitrogen and resuspended in 5 volumes (w/v) of extraction buffer (50 mM sodium borate, pH 8.5, 2 mM EDTA, 5% PVPP, 5 mM DTT, cOmplete protease inhibitor cocktail). The slurry was incubated for 20 min at 4 °C with occasional mixing. Plant debris was removed by two successive centrifugation steps at 24,000 x *g* for 30 min at 4 °C. The clear supernatant was loaded onto a 1 ml HiTrap DEAE FF column (GE Healthcare) pre-equilibrated with 20 mM sodium borate, pH 8.5, which was connected in series to an ÄKTA Prime Plus liquid chromatography system (GE Healthcare). Proteins were eluted from the column using a linear gradient from 0 M – 1 M NaCl in sodium borate, pH 8.5, over 20 ml. The flow rate was maintained at 1 ml/ min. Fractions (2 ml each) containing protein were desalted and concentrated using 5 kDa MWCO Spin-X UF 500 centrifugal concentrators until a protein concentration of 2-3 mg/ ml was reached. Concentrated protein fractions were finally combined with SDS-PAGE sample buffer. Protein concentration was routinely determined using the Qubit Fluorometer 2.0 (Life Technologies).

Fluorinase activity assay

An aliquot of the total protein extracted from the leaf, seed and flower tissue was tested for fluorination activity. The extracts were desalted and buffer-exchanged into 20 mM tris-HCl, pH 7.2, using the Zebra Spin 2 ml desalting columns. Activity assays with 20 mM tris-HCl, pH 7.2, 0.5 mM SAM, 20 mM NaF and 0.5 mg/ ml plant protein extract were set up in a total reaction volume of 200 µl. The assay was incubated at 37 °C for 3 hours. Reactions were quenched by heating at 98 °C for 10 min followed by placing on ice for 1 min. Precipitated protein was removed by centrifugation (14 000 x *g*, 10 min). An aliquot of the clear supernatant was equilibrated with 5% acetonitrile and 0.1% formic acid; and thereafter assessed for the production of 5'-FDA using LC-MS/MS.

Reaction products (25 µl injections) were separated using a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific) equipped with an Ascentis C8 guard column (5 µm, 2 cm x 4.0 mm) and an Ascentis C8 separation column (3 µm, 15 cm x 4.6 mm). The gradient

applied for separation is as follows: 5% B for 2 min, 5 – 95% B over 10 min, 95% B for 1 min, 95 – 5% B over 2 min and 5% B for 1 min; where A is 0.1 % formic acid in water and B is 0.1% formic acid in acetonitrile. The flow rate was maintained at 0.5 ml/ min. Eluted compounds were directly fed into a Bruker Compact triple quadrupole mass spectrometer where compounds were ionised (positive ion mode) with an electrospray ionisation source (ESI). The mass spectrometer was operated in the multiple reaction monitoring (MRM) mode with the following ion source parameters: collision energy: 32 V, voltage: 4500, ion spray source temperature: 220 °C and scanning range: *m/z* 50 – 1500. A MRM transition of 5'-FDA from *m/z* 270.0 to the fragment ion *m/z* 136.0 was tracked. Synthetic 5'-FDA was used as a standard. LC-MS/MS data were analysed using the Bruker Compass Data Analysis software version 4.3.

SDS-PAGE

Sodium dodecyl sulfate-polyacrylamide gels (3.2% stacker, 8% separating) were prepared according to Schägger [13]. Gels were allowed to polymerise for 12 hours prior to use to prevent free acrylamide from forming adducts with electrophoresed proteins. For broad proteomes, approximately 50 µg of total protein extracted from the seed, flower and leaf were loaded. While, for enriched proteomes, approximately 30 µg of protein from the desalted/ concentrated DEAE-fractions were loaded. Samples were electrophoresed about 0.5 cm into the separating gel at 50 V. This was conducted in order to remove salts and some contaminating small molecules that may interfere with downstream analyses. A duplicate gel, for visualisation purposes only, was performed in parallel but was allowed to electrophorese to the bottom of the gel. The gels were then stained for 2 hours with a solution of 0.025% Coomassie blue-G250, 10% acetic acid and 50% methanol. Once visible, the single protein-containing band (~0.5 cm) was excised and destained with several changes of 10% acetic acid. The destaining solution was removed and bands were stored in 10% ethanol at 4 °C until further analysis.

Peptide preparation and LC-MS/MS

Following removal from storage ethanol, excised protein bands were reduced, alkylated and digested. Firstly, gel bands were further destained using 100 mM ammonium bicarbonate in 50% acetonitrile followed by 100% acetonitrile for 5 min each at 37 °C with shaking at 550 rpm. Reduction and alkylation were then performed. Gel bands were equilibrated in 50 mM tris-HCl, pH 8.5, followed by 100% acetonitrile for 5 min each at 37 °C with shaking at 550 rpm. Bands were submerged in the reduction and alkylation buffer (50 mM tris-HCl, pH 8.5, 10 mM TCEP, 40 mM chloroacetamide) for 10 min at 95 °C with shaking at 550 rpm. This was followed by sequential washes in 100% acetonitrile, 100 mM ammonium bicarbonate and then 100% acetonitrile again. Each wash was conducted for 5 min at 37 °C with shaking at 550 rpm. Gel pieces were dried in a speedy-vac for approximately 15 min at 40 °C. For tryptic digests, gel pieces were re-swelled on ice with ~15 µl of digestion buffer (0.0125 µg/ µl modified trypsin, 40 mM ammonium bicarbonate, 5 mM $CaCl_2$) and covered with incubation buffer (40 mM ammonium bicarbonate, 5 mM $CaCl_2$). This was incubated overnight at 37 °C with shaking at 550 rpm. Peptides were recovered, dried and finally resuspended in 0.3% formic acid in water.

Resuspended peptides were subjected to nano-HPLC (50 µl injections) coupled to mass spectrometry. Peptides were separated using a Dionex UltiMate 3000 nano-HPLC system (Thermo Fisher Scientific) equipped with a C18, 5 µm, 100 Å, 5 mm x 0.3 mm enrichment column and an Acclaim PepMap RSLC nano-column (C18, 2 µm, 100 Å, 500 mm x 0.075 mm). Samples were concentrated on the enrichment column for 6 min using an isocratic flow of 0.1% formic acid at 5 µl/ min. Separation was then carried out on the nano-column at a flow rate of 250 nl/ min at 60 °C. The following gradient was used: 4% B for 6 min, 4-25% B over 88 min, 25-95% B over 5 min, 95% B for 10 min and 4% B for 15 min; where A is 0.1% formic acid in water and B is 0.1% formic acid in acetonitrile. Eluted peptides were ionised with a nanospray source equipped with stainless steel emitters (Thermo Fisher Scientific) and detected in an Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific). The MS was operated in the positive ion mode, applying an alternating full scan (m/ z 400-2000) in the ion cyclotron and MS/MS by collision-induced dissociation of the 20 most intense peaks with dynamic exclusion enabled.

Data analysis

The LC-MS/MS data were analysed by searching the proteome against the translated (6 frames) *D. cymosum* transcriptome and all common contaminants using Proteome Discoverer version 1.4 and Mascot version 2.4.1. Hits were extracted from the translated transcriptome and annotated using the UniProt database. Gene ontology (GO) enrichment was performed using REVIGO [14].

Results

Activity assay of plant protein extracts

Activity assays were conducted using the total protein extracts from the *D. cymosum* flower, leaf and seed tissue to assess for fluorination activity. The conversion of SAM into 5'-FDA represents the first committed step of the fluorination pathway, as found in numerous bacterial species. The reaction products were resolved using LC-MS/MS and monitored for the production of 5'-FDA. However, activity could not be detected in any of the assays.

Overview of the *D. cymosum* proteome

There are currently no reports detailing the proteome of *D. cymosum* therefore, despite the inability to detect fluorination activity, the proteomes were still characterised. Total protein from the leaf, seed and flower tissue; as well as protein fractions obtained following ion-exchange chromatography, were analysed using SDS-PAGE. Although the samples for shotgun sequencing were only electrophoresed 0.5 cm into the gel (to remove small contaminating molecules), a replicate gel was performed in parallel to allow for the visualisation of proteins (Figure 1).

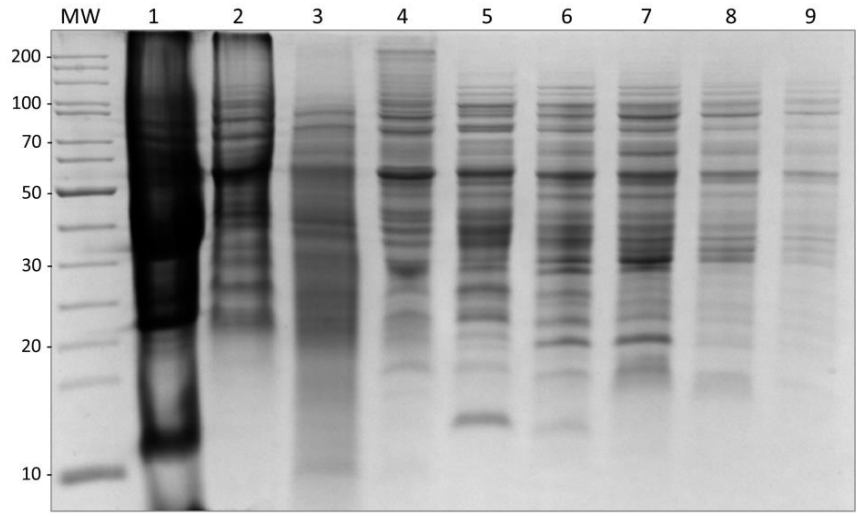


Figure 1: SDS-PAGE (8%) of the total protein extracted from the seed (lane 1), flower (lane 2) and leaf (lane 3) tissue of *D. cymosum*. Total protein extracted from the leaf of *D. cymosum* was fractionated using ion-exchange chromatography. The flow-through (lane 4) and fractions (lanes 5-9) collected were resolved.

Following peptide sequencing and data analysis; a total of 637 proteins were detected in the flower, 517 proteins were detected in the seed and 482 proteins were detected in the leaf. Of these, 179 were unique to the flower, 111 were unique to the seed and 97 were unique to the leaf (Figure 2A). Proteins extracted from the leaf tissue were further fractionated using ion-exchange chromatography in order to enrich low abundance proteins. This effectively increased the total number of proteins detected in the leaf from 482 to 1,276. In comparison, this left the flower with 90 unique proteins and the seed with 57 unique proteins (Figure 2B). Collectively, a total of 2,430 proteins were detected in *D. cymosum*.

Annotation of the *D. cymosum* proteome

The proteins identified from the total leaf extract were compiled together with proteins detected following fractionation. Collectively, 1,277 proteins were identified. From these, 1,226 were annotated with UniProt identities. A total of 7,008 gene ontologies were assigned and allocated to the biological process, molecular function and cellular component categories. GO terms were then enriched using REVIGO. Enrichment allows for functional profiling by finding a subset of representative terms to describe underlying mechanisms

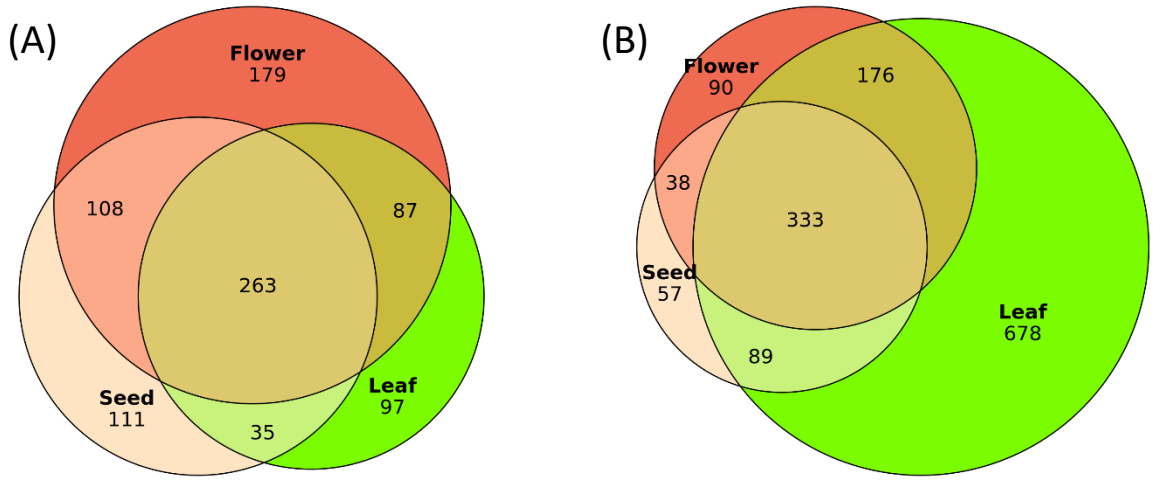


Figure 2: (A) Venn diagram showing the number of proteins detected in the flower, seed and leaf extracts of *D. cymosum*. (B) Venn diagram showing the increase in the number of proteins detected in the leaf of *D. cymosum* following fractionation using ion-exchange chromatography.

These terms are represented on a scatter plot where the log count of GO terms for each subcategory is shown on the y-axis (as well as bubble colour) and semantic differences between GO terms is shown on the x-axis. For biological process, 'carbohydrate metabolism' was the most represented process. 'Actin binding' was the most represented process for the molecular function category and 'cell' was the most represented process for cellular component (Figure 3).

For the seed, 517 proteins were detected, of which, 506 were annotated with UniProt identities. In this instance, 2,430 gene ontologies were assigned. Following GO enrichment for the seed, 'carbohydrate metabolism', 'catalytic activity' and 'cell' were the most represented terms for the biological process, molecular function and cellular component categories, respectively (Figure 4).

A total of 637 proteins were detected in the flower of *D. cymosum*. Of these, 632 were annotated with UniProt identities and 3,213 gene ontology terms were ultimately as assigned. Following GO enrichment for the flower, 'carbohydrate metabolism', 'oxidoreductase activity' and 'cell' were the most represented terms for the biological process, molecular function and cellular component categories, respectively (Figure 5).

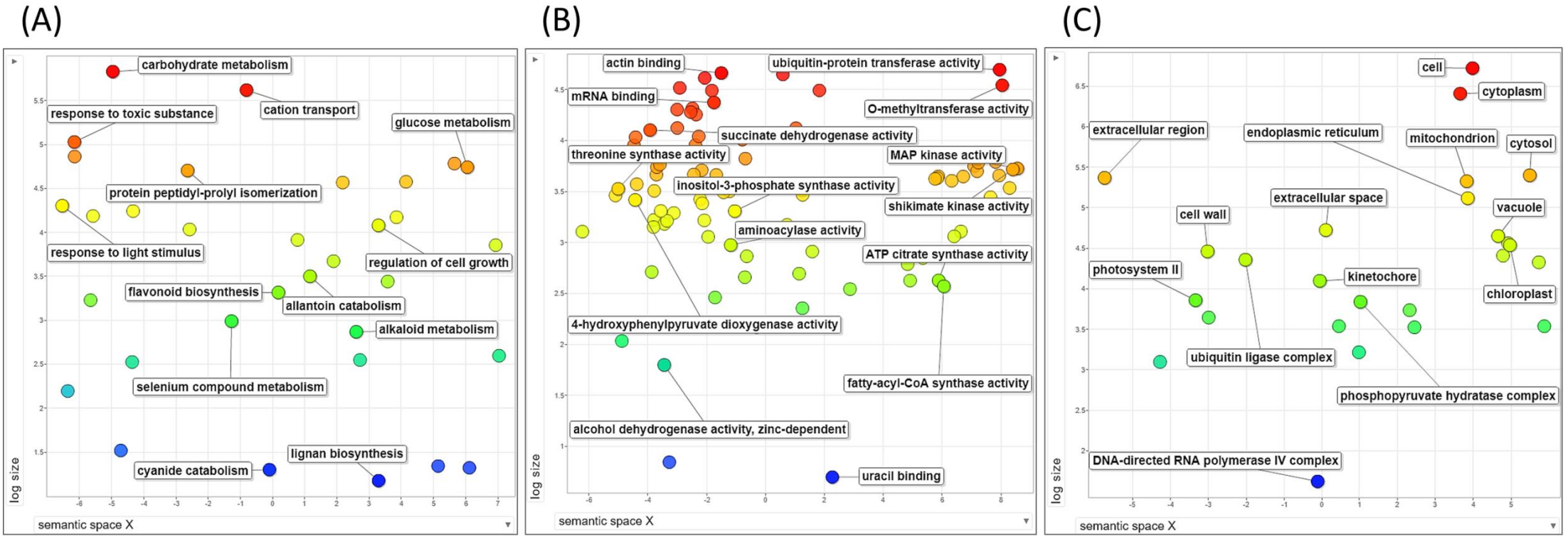


Figure 3: Enriched gene ontologies within the (A) biological process, (B) molecular function and (C) cellular component categories for proteins detected in the leaf tissue of *D. cymosum*. The log count is shown on the y-axis. Semantic space is represented on the x-axis.

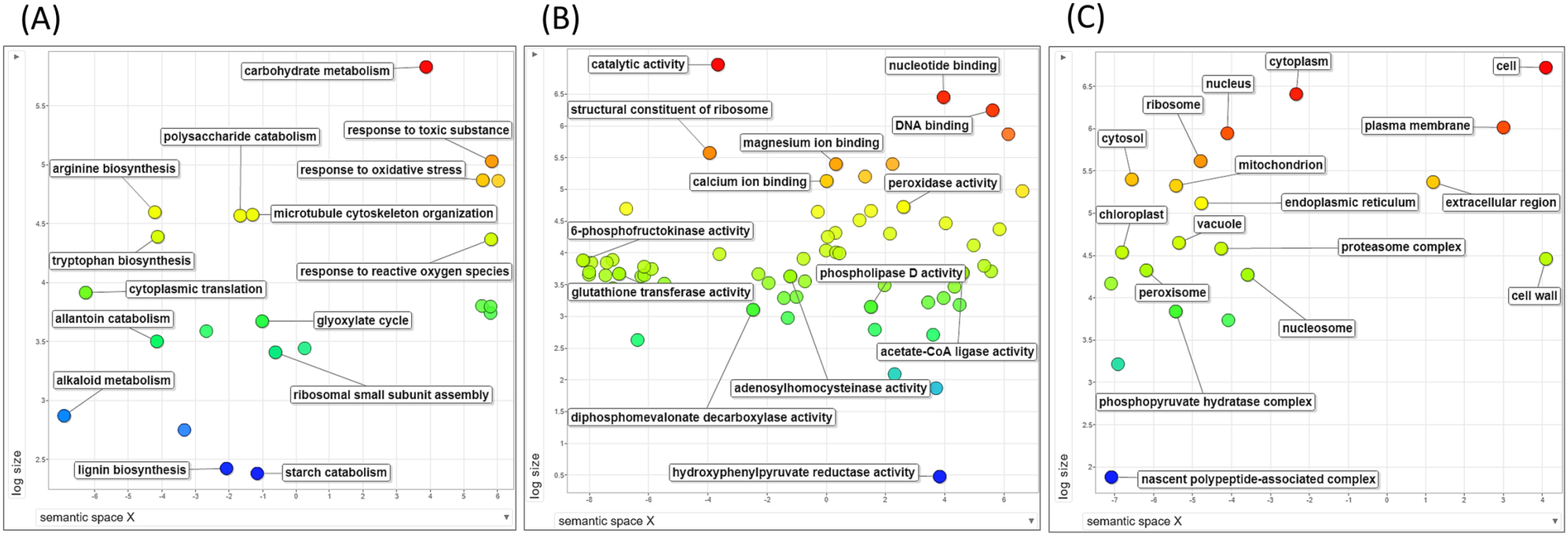


Figure 4: Enriched gene ontologies within the (A) biological process, (B) molecular function and (C) cellular component categories for proteins detected in the seed tissue of *D. cymosum*. The log count is shown on the y-axis. Semantic space is represented on the x-axis.

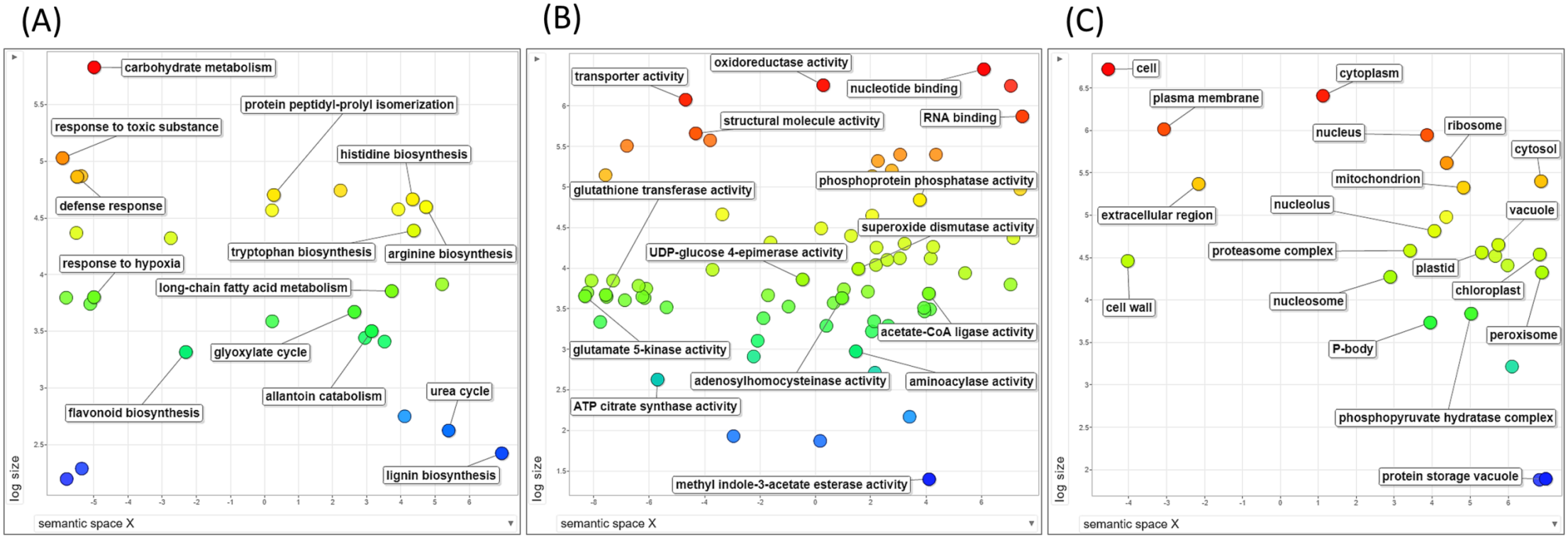


Figure 5: Enriched gene ontologies within the (A) biological process, (B) molecular function and (C) cellular component categories for proteins detected in the flower tissue of *D. cymosum*. The log count is shown on the y-axis. Semantic space is represented on the x-axis.

Discussion

Sequence similarity based approaches have failed to identify a fluorinating enzyme from *D. cymosum*. Therefore, omic information coupled with a functional screening approach was employed in this study. Previous studies have successfully used this method for enzyme discovery [11,15]. A fundamental prerequisite for this approach, however, is having a suitable activity assay. This allows for the protein of interest to be tracked through activity [10]. The total protein extracted from *D. cymosum* leaf, seed and flower tissue did not exhibit any fluorinase activity. This could be due to low quantities of the target enzyme in the assay or, perhaps, inappropriate assay conditions. It is also possible that the product, 5'-FDA, may have been consumed by the next enzyme in the pathway. Since the LC-MS/MS protocol was optimised and limited to the detection of 5'-FDA, we were unable to identify any other peaks that may relate to the intermediates that follow 5'-FDA in the fluorometabolite pathway. However, these observations would need to be confirmed through the repetition of activity assays.

Despite the inability to detect fluorinase activity, the proteome for *D. cymosum* was sequenced and characterised due to the lack of molecular information available for this plant. The workflow adopted for protein discovery in this study included analysis of the raw total protein extracts (leaf, seed and flower), as well as the analysis of protein fractions following ion-exchange chromatography (leaf only). Collectively, 1,636 proteins were detected from the total protein extractions. However, fractionation allowed for the recognition of additional proteins increasing the total number of proteins detected to 2,430. Protein fractionation assists in reducing the complexity of samples. Moreover, it helps in improving proteome coverage and detection of low abundant proteins [8]. Similar results have been reported with shotgun proteomics of the tomato [8]. The fractionation of total protein prior to proteomic analysis positively contributes to protein detection and should be expanded to the seed and flower tissue of *D. cymosum* in future studies.

Integration of proteomic data with species-specific information, such as the genome or transcriptome, frequently enhances protein detection [10]. Nynca et al reported that the lack of sequence information for trout severely limited proteomic analysis resulting in the detection of only 54 proteins [16]. In the present study, the transcriptome created for *D. cymosum* (Chapter 2) was used as a customised database for proteomic analysis. Using the same approach, a total of 2,754 proteins were detected for mango [17], paralleling that achieved here. However, an inherent limitation of this study originates from the fact that the

D. cymosum transcriptome was created exclusively from leaf tissue. This results the loss of tissue-specific information when analysing the seed or flower proteomes. This can be avoided using genomic data or tissue-specific transcriptome data in future studies.

Conclusion

In the present study, an integrative omics approach coupled with functional screening was used to investigate potential fluorinase activity in *D. cymosum*. The inability to detect fluorinase activity impeded tracking of the enzyme for identification. Nevertheless, the proteome for *D. cymosum* was characterised. Our findings showed that fractionation of total protein prior to proteomic analysis significantly increased the number of proteins detected. This work provides the first comprehensive proteome for *D. cymosum* and can be used as a platform for identification of proteins in future studies.

References

1. Marais, J. S. C. (1944) Monofluoroacetic acid, the toxic principle of "gifblaar", *Dichapetalum cymosum* (Hook) Engl. *Onderstepoort Journal of Veterinary Science and Animal Industry, 20*, 67-73.
2. Lee, S. T., Cook, D., Pfister, J. A., Allen, J. G., Colegate, S. M., Riet-Correa, F. and Taylor, C. M. (2014) Monofluoroacetate-containing plants that are potentially toxic to livestock. *Journal of Agricultural and Food Chemistry*, *62*, 7345-7354.
3. Meyer, J. M. and Grobbelaar, N. (1990) The determination, uptake and transport of fluoroacetate in *Dichapetalum cymosum*. *Journal of Plant Physiology, 135*, 546-549.
4. Grobbelaar, N. and Meyer, J. M. (1990) Fluoroacetate production by *Dichapetalum cymosum*. *Journal of Plant Physiology*, *135*, 550-553.
5. O'Hagan, D., Schaffrath, C., Cobb, S. L., Hamilton, J. T. and Murphy, C. D. (2002) Biochemistry: biosynthesis of an organofluorine molecule. *Nature, 416*, 279.
6. Deng, H., O'Hagan, D. and Schaffrath, C. (2004) Fluorometabolite biosynthesis and the fluorinase from *Streptomyces cattleya*. *Natural Product Reports*, *21*, 773-784.
7. Zhu, X., Robinson, D. A., McEwan, A. R., O'Hagan, D. and Naismith, J. H. (2007). Mechanism of Enzymatic Fluorination in *Streptomyces cattleya*. *Journal of the American Chemical Society*, *129*, 14597-14604.
8. Kilambi, H. V., Manda, K., Sanivarapu, H., Maurya, V. K., Sharma, R. and Sreelakshmi, Y. (2016) Shotgun proteomics of tomato fruits: evaluation, optimization and validation of sample preparation methods and mass spectrometric parameters. *Frontiers in Plant Science*, *7*, 969.
9. Wienkoop, S. and Weckwerth, W. (2006) Relative and absolute quantitative shotgun proteomics: targeting low-abundance proteins in *Arabidopsis thaliana*. *Journal of Experimental Botany*, *57*, 1529-1535.
10. Sturmberger, L., Wallace, P. W., Glieder, A. and Birner-Gruenberger, R. (2016) Synergism of proteomics and mRNA sequencing for enzyme discovery. *Journal of Biotechnology*, *235*, 132-138.
11. Lanfranchi, E., Pavkov-Keller, T., Koehler, E.M., Diepold, M., Steiner, K., Darnhofer, B., Hartler, J., Van Den Bergh, T., Joosten, H.J., Gruber-Khadjawi, M., Thallinger, G.G., Birner-Gruenberger, R., Gruber, K., Winkler, M. and Glieder, A. (2017) Enzyme discovery beyond homology: a unique hydroxynitrile lyase in the Bet v1 superfamily. *Scientific Reports*, *7*, 46738.

12. Vincent, D., Wheatley, M.D. and Cramer, G.R. (2006) Optimization of protein extraction and solubilization for mature grape berry clusters. *Electrophoresis*, *27*, 1853-1865.
13. Schägger, H. (2006) Tricine–SDS-PAGE. *Nature Protocols*, *1*, 16.
14. Supek, F., Bošnjak, M., Škunca, N. and Šmuc, T. (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS One*, *6*, e21800.
15. Schilmiller, A. L., Miner, D. P., Larson, M., McDowell, E., Gang, D. R., Wilkerson, C. and Last, R. L. (2010) Studies of a biochemical factory: tomato trichome deep expressed sequence tag sequencing and proteomics. *Plant Physiology*, *153*, 1212-1223.
16. Nynca, J., Arnold, G. J., Fröhlich, T. and Ciereszko, A. (2015). Shotgun proteomics of rainbow trout ovarian fluid. *Reproduction, Fertility and Development*, *27*, 504-512.
17. Wu, H.X., Jia, H.M., Ma, X.W., Wang, S.B., Yao, Q.S., Xu, W.T., Zhou, Y.G., Gao, Z.S. and Zhan, R.L. (2014) Transcriptome and proteomic analysis of mango (*Mangifera indica* Linn) fruits. *Journal of Proteomics*, *105*, 19-30.

Chapter 4

Identification and characterisation of a novel fluorinase from *Actinopolyspora mzabensis*

Selisha A. Sooklal, Charles de Koning, Dean Brady and Karl Rumbold

This chapter has been submitted for publication.
Protein Expression and Purification (Under review)

Identification and characterisation of a novel fluorinase from *Actinopolyspora mzabensis*

Selisha Ann Sooklal[a], Charles de Koning[b], Dean Brady[b] and Karl Rumbold[a*]

[a] Department of Microbiology, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, 2000, South Africa

[b] School of Chemistry, University of the Witwatersrand, Johannesburg, 2000, South Africa

[*] Corresponding author: Department of Microbiology, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, 2000, South Africa. E-mail: karl.rumbold@wits.ac.za

Abstract

The incorporation of fluorine has been shown to improve the biophysical and bioactive properties of several organic compounds. However, sustainable strategies of fluorination are needed. Fluorinases have the unique ability to catalyse a C-F bond, hence, have vast potential to be applied as biocatalysts in the preparation of fine chemicals. But fluorinases are extremely rare in nature with only five representatives isolated thus far. Moreover, the heterologous expression of fluorinases is challenged by low yields of soluble protein. This study describes the identification of a novel fluorinase from *Actinopolyspora mzabensis*. Overexpression of the *Am*-fluorinase in *E. coli* BL21 (DE3) resulted in the formation of inclusion bodies (IBs). The enzyme was recovered from IBs, solubilised in 8 M urea, and successfully refolded into a biologically active form. Following hydrophobic interaction chromatography, >80 mg of the active fluorinase was obtained at a purity suitable for biocatalytic applications. An additional gel filtration step gave ≥ 95% pure *Am*-fluorinase. Using LC-MS/MS, the optimal pH for activity was found at 7.2 while the optimal temperature was about 65°C. At these conditions, the enzyme exhibited an activity of 0.44 U/mg. Furthermore, the *Am*-fluorinase showed exceptional stability at 25°C. Preliminary results suggest that the newly identified *Am*-fluorinase is thermostable.

Keywords: Fluorinase, *Actinopolyspora mzabensis*, inclusion bodies, protein refolding, C-F bond, biocatalyst

Abbreviations used: IBs, inclusion bodies; SAM, *S*-adenosyl-$_L$-methionine; 5'-FDA, 5'-fluoro-5'-deoxyadenosine.

1. Introduction

The incorporation of fluorine into organic molecules has been shown to improve the biophysical and bioactive properties of compounds commonly used in the pharmaceutical, agrochemical and material industries. In effect, it has been estimated that 25% of drugs on the market contain fluorine [1]. And this number is about the same for agrochemicals underscoring the commercial significance of fluorine [2]. The traditional approach for fluorination of organic molecules, however, has been through chemical synthesis. But due to the electronegativity and high hydration energy of the fluoride ion, it renders the chemical incorporation of this atom challenging which often requires harsh conditions and lacks selectivity [3,4]. Therefore, innovative and sustainable fluorination strategies are needed.

In 2002, O'Hagan et al reported the discovery of the first fluorinating enzyme, or fluorinase, from the bacterium *Streptomyces cattleya* [5,6]. This finding revolutionised the field of fluorination by demonstrating the enzyme catalysed formation of a C-F bond. Fluorinases mediate the conversion of *S*-adenosyl-$_L$-methionine (SAM) and inorganic fluoride ions into 5'-fluoro-5'-deoxyadenosine (5'-FDA) and $_L$-methionine [6,7]. This represents the first step in the production of fluorinated metabolites. The pathway then proceeds with four additional enzymatic steps to yield either fluoroacetate (FA) or 4-fluorothreonine (4-FT) [8], as shown in Fig. 1. More recently, another fluorinated metabolite 5'-fluoro-2,3,4-trihydroxypentanoic acid (5'-FHPA) was found exclusively in *Streptomyces* sp. MA37 leading to a branch in this pathway [9].

Enzyme-based catalysis, or biocatalysis, addresses the limitations associated with chemical synthesis by offering highly selective, safe and sustainable biotransformation [10]. With the exclusive ability to catalyse a C-F bond, fluorinases are an attractive alternative for the preparation of fine chemicals. Accordingly, these enzymes have considerable potential in industry [8] and have most prominently been applied in the synthesis of [^{18}F]-radiolabelled

products for Positron Emission Tomography (PET) [11-16] as well as synthetic biology approaches to produce a variety of fluorinated metabolites [17,18]. Despite recent advances in sequencing technologies though, fluorinases are still extremely rare and only four additional representatives have been isolated from *Streptomyces* sp. MA37, *Nocardia brasiliensis*, *Actinoplanes* sp. N902-109 [19,20] and *Streptomyces xinghaiensis* [21]. This has not only created a lack in the library of fluorinating enzymes, but has also restricted the range of biocatalytic activities available to exploit for industrial use.

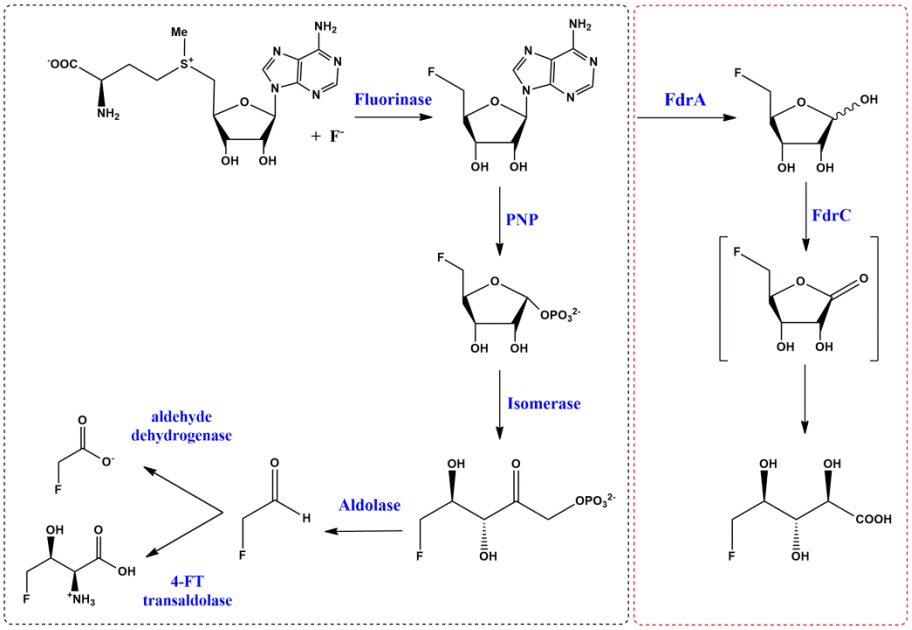


Fig. 1: Biosynthetic pathway for fluorometabolites. The pathway for the production of fluoroacetate and 4-fluorothreonine is shown in the black box. A branch in this pathway (red box) was found in *Streptomyces* MA37 which leads to the production 5‘-fluoro-2,3,4-trihydroxypentanoic acid.

Apart from the scarcity of these enzymes, the in-depth application of fluorinases is also challenged by recovery yields following heterologous expression [20]. While production of the enzyme in *Escherichia coli* has provided a convenient expression system, obtaining ample quantities of the protein is problematic. This is because majority of the recombinant fluorinase forms insoluble aggregates or inclusion bodies (IBs) in bacterial cells [19,21]. Until better expression systems are explored, it is important to find methods for the efficient recovery of the enzyme.

The formation of IBs offers some selective advantages over soluble protein such as (i) simple isolation from lysed cells (ii) resistance to proteolytic attack and (iii) high-levels of recombinant expression together with fewer contaminating host cell proteins facilitates subsequent purification [22]. Biologically active protein can be acquired from IBs provided that suitable strategies for the solubilisation and renaturation of the target protein into its native state are used. However, previous studies have traditionally opted to salvage the fluorinase from the soluble protein fraction [7,19-21] and the recovery of this enzyme from IBs has not been reported.

Actinopolyspora mzabensis is a halophilic actinomycete that was isolated from Saharan soil in the Mzab region, southern Algeria [23]. The genome of this organism was recently sequenced and deposited onto a public database. We report the identification and characterisation of a fluorinase from *A. mzabensis*. In addition, we describe a method for the efficient recovery of the fluorinase from inclusion bodies and subsequent protein refolding. To the best of our knowledge, this is the first report of an additional fluorinase since 2016 [21] as well as the first report on the successful refolding of the enzyme.

2. Materials and methods

2.1 Materials

The ampicillin, isopropyl β-D-1-thiogalactopyranoside (IPTG), chicken egg-white lysozyme, *S*-adenosyl-$_L$-methionine (SAM), NaF and triton-X 100 were purchased from Sigma-Aldrich (Missouri, USA). The restriction enzymes were from New England BioLabs (Massachusetts, USA) and the *E. coli* BL21 (DE3) cells were from Novagen (Madison, USA). The DNase I, PageRuler unstained protein ladder and 10 kDa MWCO SnakeSkin dialysis tubing were purchased from Thermo Fisher Scientific (Massachusetts, USA). The 30 kDa MWCO Spin-X UF 20 centrifugal concentrators were purchased from Corning (New York, USA). The BioRad mini-PROTEAN gel electrophoresis system was used for SDS-PAGE (California, USA). The HPLC-grade acetonitrile, formic acid and dithiothreitol (DTT) were from VWR (Pennsylvania, USA). The synthetic 5'-fluoro-5'-deoxyadenosine (5'-FDA) was a kind gift from Dr. Allan Prior (now at the University of Hawaii). All other general reagents were purchased from either VWR (Pennsylvania, USA) or SRL (Mumbai, India), unless otherwise stated.

2.2 Identification of the fluorinase and fluorometabolite gene cluster in *Actinopolyspora mzabensis*

A blastp search was conducted against the non-redundant database using the known fluorinase sequences as the query. A hypothetical protein (GenBank accession: **SDK90789.1**) from *Actinopolyspora mzabensis* was chosen. The gene sequence of the hypothetical protein, hereon denoted as *Am-fluorinase*, was retrieved from the genomic sequence (GenBank accession: **FNFM01000017.1**) of *A. mzabensis*. The fluorometabolite gene cluster and open reading frames (ORFs) were identified using 2ndFind [24].

2.3 Multiple sequence alignment, phylogenetic tree and structure modelling

Alignment of the fluorinase, chlorinase and hydroxide adenosyltransferase protein sequences were performed on the CLC Genomics Workbench version 12.0 (Qiagen). A bootstrapped phylogenetic tree was also generated on the CLC Genomics Workbench using the neighbor-joining method. All accession numbers and corresponding sequences used in this study are given in Table S1. Homology models were created *in silico* using MODELLER version 9.19 [25]. The fluorinase from *S. cattleya* (PDB accession: 1RQP) was employed as a template to predict five models for the *A. mzabensis* fluorinase (*Am*-fluorinase). The model with the lowest discrete optimised potential energy (DOPE) score was selected. The PyMOL Molecular Graphics System version 2.0.7 (Schrödinger) [26] was then used for visualisation and analysis of the structure.

2.4 Plasmids

A pET-11a plasmid containing the codon-optimised *Am-fluorinase* gene (pET-11a-*Am-fluorinase*) was purchased from Genscript, USA. The synthetic gene was flanked by the restriction sites *Nde*I, upstream of the ATG start codon; and *Bam*HI, downstream of the UAA stop codon. The resultant sequence did not contain any protein purification-associated tags.

2.5 Protein expression, refolding and purification

The pET-11a-*Am-fluorinase* plasmid was transformed into chemically competent *E. coli* BL21 (DE3) cells using the heat shock method [27] and plated on Luria-Bertani (LB) agar supplemented with 0.15 g/L $MgSO_4$ and 100 µg/ml ampicillin. Colony PCR (universal T7 primers) and restriction digests (*Nde*I and *Bam*HI; *Kpn*I) were used to screen for positive transformants. *E. coli* BL21 (DE3) cells containing the plasmid were grown in 2×YT broth supplemented with 0.15 g/L $MgSO_4$ and 100 µg/ml ampicillin. Cultures (1 L) were grown at 37 °C with shaking (200 rpm) until an optical density of ~ 0.6 at 600 nm was reached. Over-expression of the *Am*-fluorinase was then induced using 0.2 mM IPTG and cultures were further incubated at 25 °C with shaking (200 rpm) for 12-14 hours. Cells were harvested by centrifugation at 5000 x *g* for 15 min at 4 °C and stored at – 20 °C.

Thawed cell pellets were resuspended in lysis buffer (50 mM tris-HCl, pH 8.0, 1 mM DTT, 100 µg/ml lysozyme) and incubated for 30 min on ice. Further lysis was achieved through 6 cycles of sonication on ice at 50% duty (20 second bursts and 40 second rest periods). The cell lysate was centrifuged (24 000 x *g*, 30 min, 4 °C) to separate the soluble and insoluble protein fractions. The insoluble pellet, containing the inclusion bodies (IBs), was resuspended in wash buffer W1 (50 mM tris-HCl, pH 8.0, 1 mM $MgCl_2$, 1% tritonX-100, 5 mM DTT) and incubated at 25 °C for 30 min with 1 U/ml of DNase I to remove contaminating DNA. This was centrifuged (20 000 x *g*, 30 min, 4 °C) and the pellet was resuspended in wash buffer W2 (50 mM tris-HCl, pH 8.0, 1 mM EDTA, 1% tritonX-100, 1 M urea, 5 mM DTT). Following centrifugation (20 000 x *g*, 30 min, 4°C), the *Am*-fluorinase was recovered from inclusion bodies by resuspending the pellet in a denaturing buffer (50 mM tris-HCl, pH 8.0, 1 mM EDTA, 100 mM NaCl, 5% (v/v) glycerol, 8 M urea, 20 mM DTT). This was incubated at 4 °C for 30 min and then centrifuged at 24 000 x *g* for 30 min at 4 °C. The denatured *Am*-fluorinase, in the supernatant, was refolded into its native form through stepwise dialysis against refolding buffer (50 mM tris-HCl, pH 8.0, 100 mM NaCl, 5% (v/v) glycerol, 1 mM DTT) containing 4, 2, and 0 M urea. Dialysis was carried out using 10 kDa MWCO SnakeSkin dialysis tubing and each step was conducted for 6-8 hours at 4 °C.

Prior to purification, the refolded *Am*-fluorinase was centrifuged at 24 000 x *g* for 30 min at 4 °C to remove insoluble aggregates. The clear supernatant was applied onto 2 x 5 ml HiTrap Phenyl FF columns (GE Healthcare) pre-equilibrated with 20 mM tris-HCl, pH 7.2,

100 mM NaCl and 5% (v/v) glycerol; which was connected in series to an ÄKTA Prime Plus liquid chromatography system (GE Healthcare). Protein was eluted in a single step with 20 mM tris-HCl, pH 7.2, 5% (v/v) glycerol. The flow rate was maintained at 1.5 ml/min. Fractions containing the *Am*-fluorinase were precipitated with 1 M $(NH_4)_2SO_4$ from a tris-buffered stock solution (pH 7.5) and centrifuged (20 000 x *g*, 20 min, 4 °C). The pellet containing the fluorinase was resuspended in 50 mM tris-HCl, pH 7.2, 1 mM DTT and dialysed overnight against 50 mM tris-HCl, pH 7.2, 300 mM NaCl, 1 mM DTT at 4 °C to remove residual $(NH_4)_2SO_4$ and equilibrate the sample for size exclusion chromatography. The protein was concentrated using a 30 kDa MWCO Spin-X UF 20 centrifugal concentrator and filtered through a 0.22 µm membrane. Concentrated protein (1% column volume) was loaded onto a HiPrep 26/60 Sephacryl S-300 HR column (GE Healthcare) equilibrated with 20 mM tris-HCl, pH 7.2, 300 mM NaCl. The flow rate was maintained at 1 ml/min. Only fractions that contained ≥ 95% pure *Am*-fluorinase were pooled, concentrated and dialysed against 10 mM Tris-HCl, pH 7.2, 1 mM DTT for characterisation studies.

Protein concentrations were determined using the Qubit Fluorometer 2.0 (Invitrogen) while protein purity was routinely assessed using 14% tricine SDS-PAGE [28] and densitometry (Fiji) [29]. The theoretical molecular weight of the *Am*-fluorinase was determined using the ExPASy ProtParam tool [30].

2.6 Verification of fluorinase activity

To ascertain if the recombinant protein was indeed a fluorinase, fluorination activity was tested. An assay with 20 mM tris-HCl, pH 7.2 buffer, 20 mM NaF, 0.5 mM SAM and 2 µM purified *Am*-fluorinase in a total reaction volume of 200 µL was performed at 37 °C for 20 min. Appropriate controls containing synthetic 5'-FDA; or reactions without NaF, without SAM and without the purified *Am*-fluorinase were executed in parallel. Reactions were quenched by heating at 98 °C for 10 min followed by placing on ice for 1 min. Precipitated protein was removed by centrifugation (14 000 x *g*, 10 min). An aliquot of the clear supernatant was equilibrated with 5% acetonitrile and 0.1% formic acid; and thereafter assessed for the production of 5'-FDA using LC-MS/MS.

2.7 Effect of pH, temperature and metal ions on fluorinase activity

The optimal pH value for the fluorinase from *A. mzabensis* was determined using individual assays (200 µL) containing 20 mM NaF and 0.5 mM SAM in 20 mM sodium acetate (pH 5.2), sodium phosphate (pH 6.0) or tris-HCl (pH 7.2- 8.5) buffers. The reactions were initiated through the addition of 3 µM *Am*-fluorinase and incubated at 37 °C for 20 min. To measure the effect of temperature on fluorinase activity, reactions (200 µL) with 20 mM tris-HCl, pH 7.2 buffer, 20 mM NaF and 0.5 mM SAM were initiated with 3 µM *Am*-fluorinase and incubated for 20 min at various temperatures (25 °C - 80 °C). The influence of metal ions on fluorinase activity was determined using reactions with 20 mM tris-HCl, pH 7.2 buffer, 20 mM NaF, 0.5 mM SAM and 1 mM of various divalent metal chloride salts (Mg^{2+}, Mn^{2+}, Cu^{2+}, Co^{2+}, Fe^{2+} and Zn^{2+}) or 1 mM EDTA to remove any associated metal ions. A control without metal ions or EDTA was carried out. Reactions were initiated with the addition of 3 µM *Am*-fluorinase and incubated at 37 °C for 30 min.

All reactions were quenched and equilibrated as previously described; and subjected to LC-MS/MS analysis. The production of 5'-FDA was quantified against a standard curve of synthetic 5'-FDA. Relative activity is reflected as a percentage of the specific activity of the fluorinase. All experiments were performed in triplicate and error bars indicate the standard deviation.

2.8 Protein stability at room temperature

The purified *Am*-fluorinase was stored at either 4 °C or 25 °C in 10 mM tris-HCl, pH 7.2 buffer for 0 days (control) and 2 months. Activity was assayed on the respective days in 200 µL reactions containing 20 mM tris-HCl, pH 7.2, 20 mM NaF, 0.5 mM SAM and 3 µM *Am*-fluorinase at 37 °C for 20 min. Reactions were quenched, equilibrated and subjected to LC-MS/MS for the quantification of 5'-FDA. Relative activity is reflected as a percentage of the specific activity of the fluorinase. The experiment was performed in triplicate and error bars indicate the standard deviation.

2.9 LC-MS/MS analysis

An aliquot (25 µL) of the enzyme reaction products was separated using a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific). The UHPLC was equipped with an Ascentis C8 guard column (5 µm, 2 cm x 4.0 mm) and an Ascentis C8 separation column (3 µm, 15 cm x 4.6 mm). The gradient applied for separation is as follows: 5% B for 2 min, 5 – 95% B over 10 min, 95% B for 1 min, 95 – 5% B over 2 min and 5% B for 1 min; where A is 0.1% formic acid in water and B is 0.1% formic acid in acetonitrile. The flow rate was maintained at 0.5 ml/min. Eluates were directly fed into a Bruker Compact triple quadrupole mass spectrometer where compounds were ionised (positive ion mode) with an electrospray ionisation source (ESI). The mass spectrometer was operated in the multiple reaction monitoring (MRM) mode and parameters were optimised for the selective detection of 5'-FDA using the synthetic standard. The following ion source parameters were used: collision energy: 32 V, ion spray voltage: 4500, ion spray source temperature: 220 °C and scanning range: *m/z* 50 – 1500. LC-MS/MS data were analysed using the Bruker Compass Data Analysis software version 4.3.

3. Results

3.1 Sequence analysis

A putative fluorinase was identified in the genome of the soil bacterium, *A. mzabensis*. Extending the search up-stream and down-stream of the *Am-fluorinase* gene, an ~ 8 kb fragment (7 genes) of the fluorometabolite gene cluster was found (Fig. S1). The *Am-fluorinase* gene retrieved was 897 bp in length encoding a polypeptide of 299 amino acids. The protein sequence shared 79% identity with the fluorinase from *S. cattleya*.

In comparison to the fluorinase, chlorinases and hydroxide adenosyltransferases are the most closely related enzymes which exhibit about 40% and 30% similarity, respectively [31-33]. These enzymes perform similar functions on the C-5' of SAM with nucleophiles excluding the fluoride ion [33], as depicted in Fig. S2. Alignment of these sequences reveals a 22 amino acid motif between residues K92 and A113, found exclusively in fluorinases [7,34]. This feature is seemingly conserved in the *Am*-fluorinase (Fig. 2, green box). In addition, residues (S158, F156, T80 and D16) that are crucial for catalytic activity were also highly conserved throughout all the fluorinase sequences, including the *Am*-fluorinase (Fig.

2, blue stars). The phylogenetic relationship between these enzymes (Fig. S3) showed that the *Am*-fluorinase forms a distinct group with the five known fluorinases. This group is separate from either the chlorinase or hydroxide adenosyltransferases.

A three-dimensional model was produced *in silico* to evaluate the structure of the *Am*-fluorinase monomer. The fluorinase from *S. cattleya* (PDB accession: **1RQP**) was used as a template and the model with the lowest DOPE score was selected. The model had a coverage of 97% and high similarities in structure was seen. The 22 amino acid motif formed a loop and is indicated by an arrow in Fig. S4.

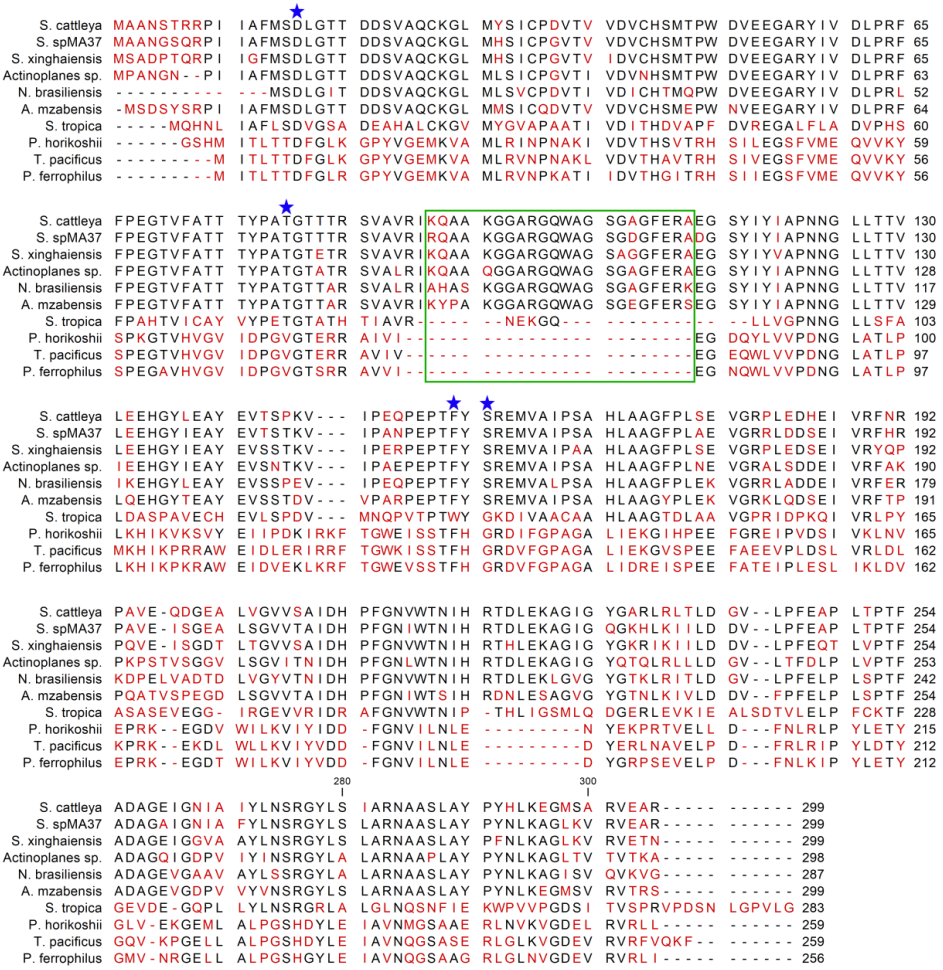


Fig. 2: Multiple sequence alignment comparing the fluorinases (*S. cattleya*, *Streptomyces* MA37, *S. xinghaiensis*, *Actinoplanes* sp., *N. brasiliensis* and *A. mzabensis*), **chlorinase** (*S. tropica*) **and hydroxide adenosyltransferases** (*P. horikoshii*, *T. pacificus* and *P. ferrophilus*). The 22 amino acid motif, a distinguishing feature of known fluorinases, is shown in the green box. Conserved residues essential for catalytic activity in fluorinases are indicated by blue stars.

3.2 Cloning, expression and refolding of the fluorinase

The codon-harmonised *Am-fluorinase* gene, encoding a putative fluorinase, was synthesized and sequenced at GenScript (USA). The pET 11a-*Am-fluorinase* plasmid was inserted into *E. coli* BL21 (DE3) cells and screened using both colony PCR and restriction digests (Fig. S5). Positive transformants were selected for overexpression, purification and characterisation of the *Am*-fluorinase.

Overexpression of the putative fluorinase in *E. coli* cells was induced with 0.2 mM IPTG. Following induction, cells were lysed and centrifuged to separate the soluble and insoluble protein fractions (Fig. 3A). While there were some detectable quantities of *Am*-fluorinase in the soluble protein fraction, it became evident that majority of the enzyme accumulated in the insoluble fraction as aggregates or inclusion bodies (IBs). This is shown by the intense protein band (~ 32 kDa) in the insoluble fraction (Fig. 3A, lane 1) which corresponds to the expected molecular weight of the *Am*-fluorinase monomer.

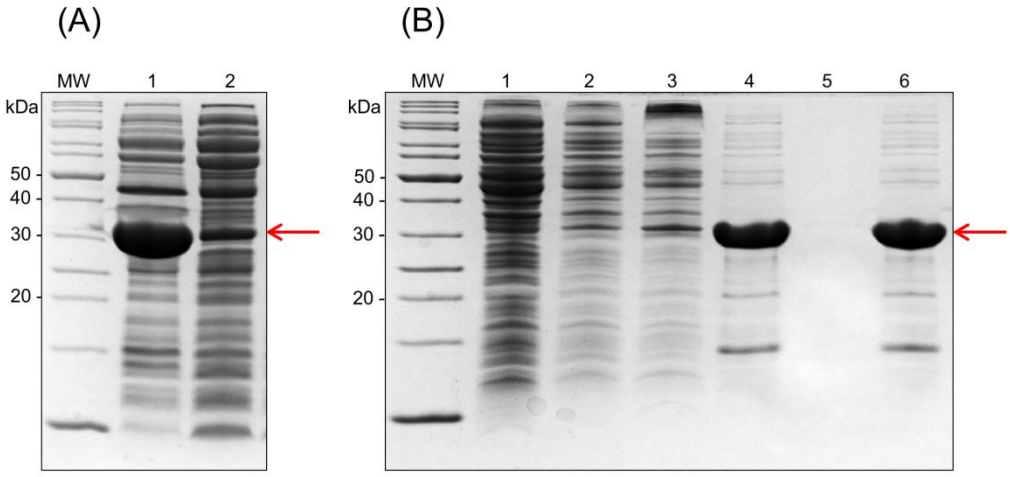


Fig. 3: SDS-PAGE of (A) insoluble (lane 1) **and soluble** (lane 2) **protein fractions. (B) Washes, denaturation and refolding of IBs.** Lane 1 is the soluble protein fraction following cell lysis and centrifugation. The pellet containing the IBs were washed in a buffer with 1 % triton-X 100 (lane 2), followed by 1 M urea (lane 3). The IBs were then denatured in a buffer containing 8 M urea and 20 mM DTT (lane 4). Following dialysis, lane 6 shows the refolded fluorinase in 0 M urea. MW refers to the molecular weight marker (PageRuler unstained protein ladder). The protein band corresponding to the fluorinase (~ 32 kDa) is indicated by the red arrow.

Inclusion bodies were harvested following cell lysis and centrifugation. The insoluble pellet contained approximately 26% *Am*-fluorinase as a fraction of the total protein (Fig. 3A, lane 1). Prior to IB solubilisation, contaminating proteins were reduced through two was steps with no significant influence on the *Am*-fluorinase (Fig. 3B). The pellet was first washed in a buffer containing 1% triton-X 100 (Fig. 3B, lane 2) and subsequently washed with a buffer containing a moderate concentration (1 M) of urea (Fig. 3B, lane 3). This resulted in the content of *Am*-fluorinase increasing to 59% as seen in Fig. 3B (lane 4) following solubilisation in 8 M urea. The denatured IBs were then refolded through stepwise dialysis in buffers gradually decreasing the concentration of urea. Fig. 3B, lane 6 shows the *Am*-fluorinase in 0 M urea.

3.3 Purification of the fluorinase

The recombinant *Am*-fluorinase did not contain any protein-purification associated tags. Hydrophobic interaction chromatography (HIC) was, therefore, optimised. Prior to HIC, samples are commonly equilibrated with up to 1.5 M $(NH_4)_2SO_4$ to facilitate the binding of proteins to the column [6]. The use of $(NH_4)_2SO_4$ resulted in the precipitation of the *Am*-fluorinase. Hence, NaCl, a weaker salt on the Hofmeister series was investigated. The binding conditions were optimised using NaCl at 100 mM, 200 mM and 400 mM (Fig. S6). The refolded protein sample was able to bind to the column at all NaCl concentrations with no apparent precipitation during the preceding equilibration step. At 100 mM NaCl, however, the target protein was able to bind more selectively as opposed to contaminants (Fig. S6). Moreover, the refolded protein sample could be loaded directly onto the column as 100 mM NaCl was included in the refolding buffer thereby eliminating the need for pre-equilibration. Bound protein was eluted in a single step with a buffer containing 0 M NaCl. The purification resulted in fractions containing up to 80% pure *Am*-fluorinase (Fig. 4A). Using the procedure developed here, > 80 mg of fluorinase per litre of culture can be recovered following HIC.

Although the aforementioned purity of the fluorinase may be considered adequate for many applications, an additional purification step was performed to obtain a purer preparation of the *Am*-fluorinase for characterisation purposes. Fractions, following HIC, were precipitated with 1 M $(NH_4)_2SO_4$ and resuspended in a small volume of buffer in order to concentrate the sample. Dialysis was then performed to equilibrate the sample for size exclusion

chromatography (SEC). Fig. 4B shows that ≥ 95% pure *Am*-fluorinase was obtained following SEC. Fractions containing the *Am*-fluorinase were concentrated, dialysed and used for preliminary characterisation studies.

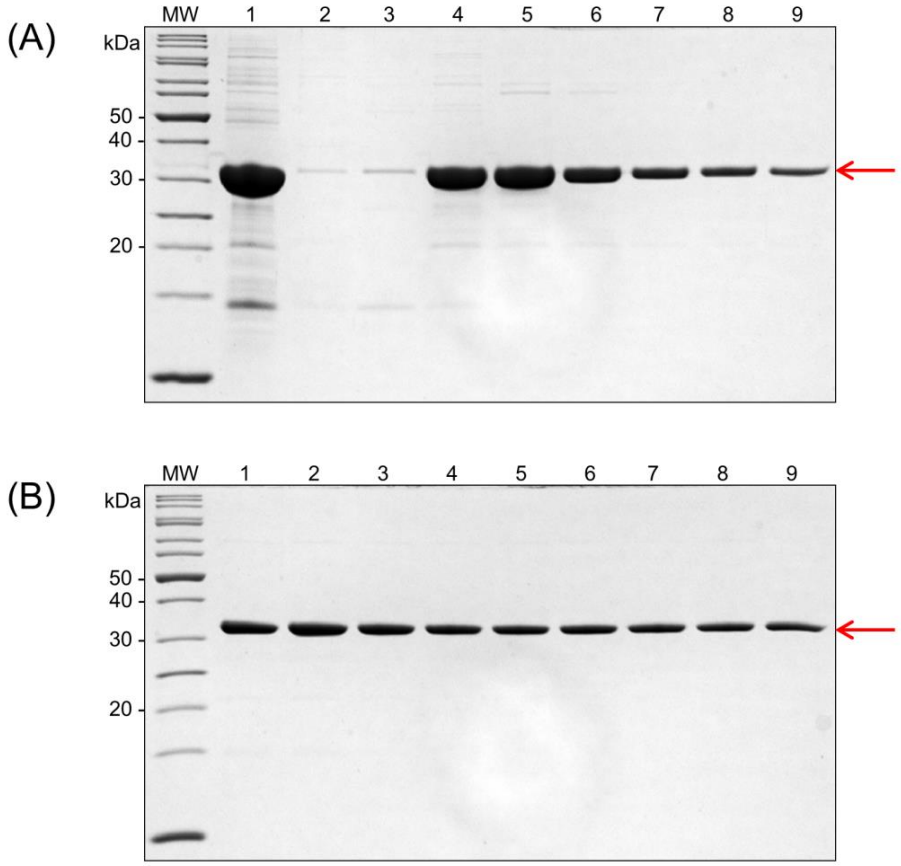


Fig. 4: Purification of the refolded fluorinase following (A) hydrophobic interaction chromatography. The refolded fluorinase was loaded (lane 1) onto a phenyl sepharose column. The protein was eluted in a single step and fractions (lanes 2- 9) were collected. **(B) Size exclusion chromatography.** Lane 1- 9 shows the fractions containing ≥ 95 % pure fluorinase collected following SEC. MW refers to the molecular weight marker (PageRuler unstained protein ladder). The protein band corresponding to the fluorinase (~ 32 kDa) is indicated by the red arrow.

3.4 Verification of a novel fluorinase

Fluorinases catalyse a C-F bond by converting SAM and inorganic fluoride ions to 5'-FDA (Fig. 5A). Using LC-MS/MS for the selective detection of 5'-FDA, fluorination activity of

the crude cell lysate from *E. coli* and the purified recombinant protein were monitored. Synthetic 5'-FDA as well as the enzymatic production of 5'-FDA using the fluorinase from *S. cattleya* (retention time: 6 min; *m/z* 270.0 → *m/z* 136.0 for both) were used as positive controls. When assayed, both the crude cell lysate of *E. coli* harbouring the pET-11a-*Am-fluorinase* plasmid (Fig. S7) and the purified *Am*-fluorinase (Fig. 5 B and C) produced a peak (retention time: 6 min; *m/z* 270.0 → *m/z* 136.0) consistent with that of 5'-FDA. The reaction could not occur in the absence of either substrate, SAM or NaF, or in the absence of the recombinant *Am*-fluorinase.

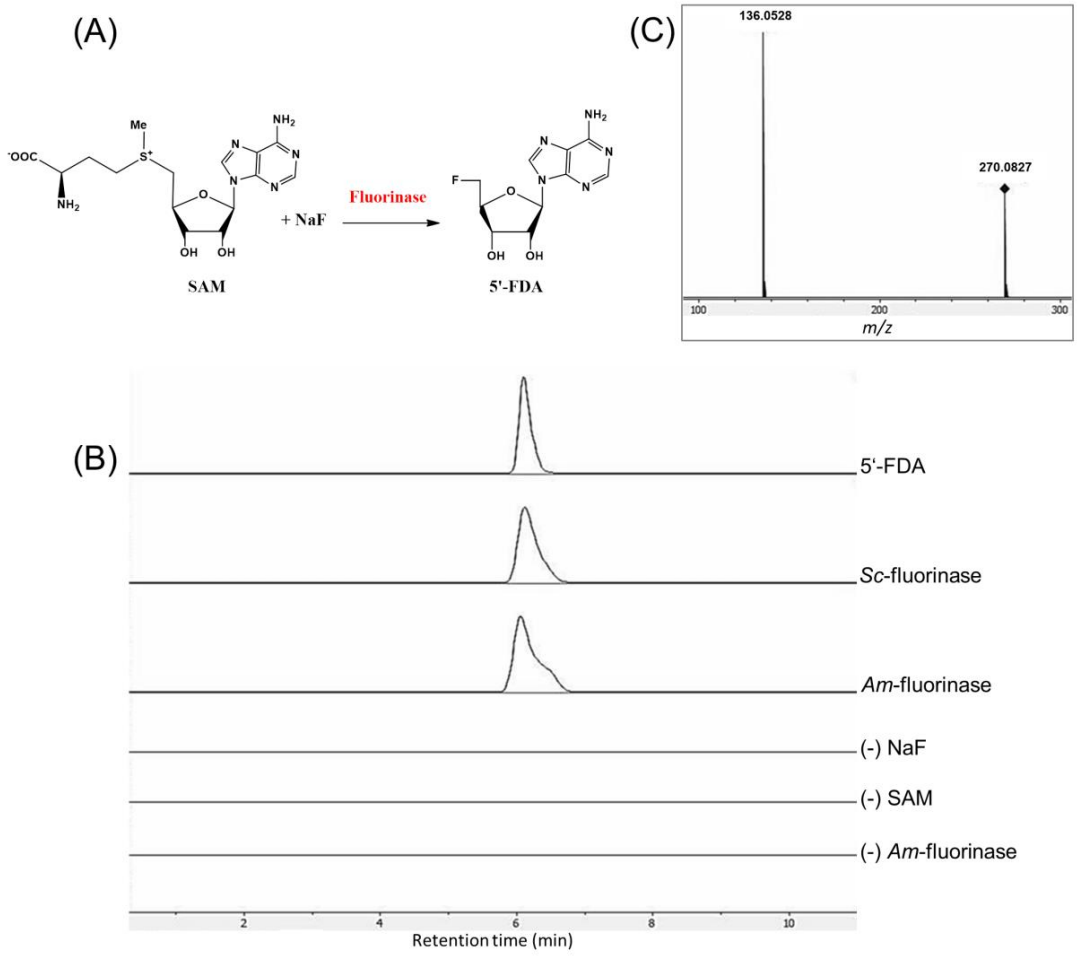


Fig. 5: Verification of fluorinase activity. (A) The reaction catalysed by the fluorinases **(B)** LC-MS/MS analysis of synthetic 5'-FDA as well as the reaction products of fluorinases from *S. cattleya* (control) and *A. mzabensis*. These produced a peak with a retention time of 6 min. Control reactions excluding NaF, SAM or the *Am*-fluorinase did not produce 5'-FDA **(C)** The corresponding mass spectrum for the peak produced in (B) showing the MRM transition of 5'-FDA (*m/z* 270.0) into the fragment ion (*m/z* 136.0).

3.5 Functional characterisation

3.5.1 Effect of pH and temperature

The optimal pH for activity of the fluorinase from *A. mzabensis* was investigated using various buffers at a pH range of 5.2 - 8.2. The production of 5'-FDA was quantified based on a standard curve using LC-MS/MS (Fig. S8). Maximum activity was observed in a tris-HCl buffer at pH 7.2 (Fig. 6A). The activity rapidly decreased at more alkaline pH values. However, activity was not as severely affected by slightly acidic values with 62% of activity retained at pH 6. Similarly, the effect of temperature on the *Am*-fluorinase activity was explored between 25 °C and 80 °C. The optimal temperature for activity was around 65 °C (Fig. 6B). Notably, the *Am*-fluorinase showed exceptional stability at higher temperatures retaining > 70% of activity at 80 °C. At optimal conditions, the fluorinase had an activity of 0.44 ± 0.03 U/ mg.

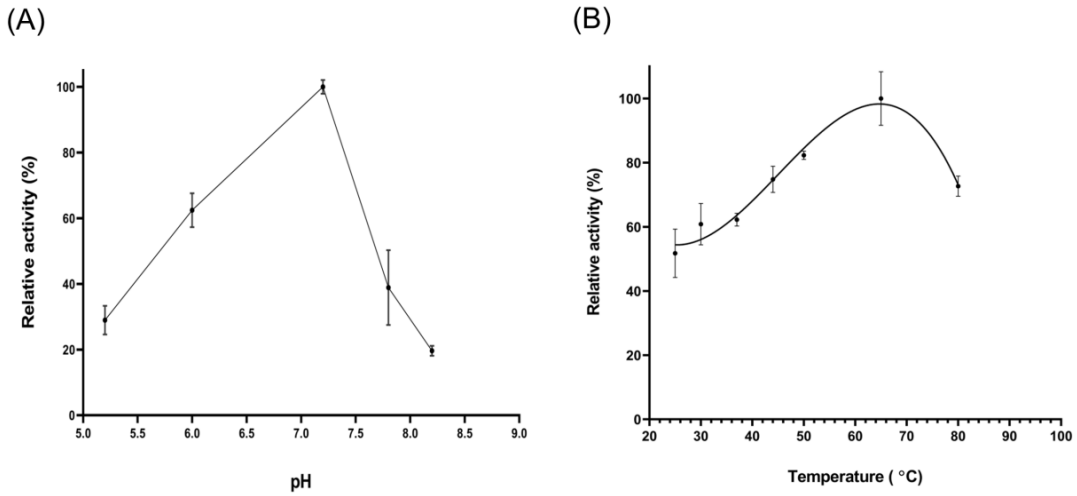


Fig. 6: Enzymatic activity as a function of (A) pH and (B) temperature for the fluorinase from *A. mzabensis*. Maximum activity observed at pH 7.2 (A) and at a temperature of 65 °C were individually set at 100 % activity. Error bars represent the standard deviation obtained from a minimum of 3 replicates.

3.5.2 Influence of divalent metal ions

The influence of various divalent metal ions at a concentration of 1 mM was examined. Fluorination activity was not severely affected by most metal ions (Mg^{2+}, Mn^{2+}, Co^{2+} and Fe^{2+}), as shown in Fig. 7. However, both Cu^{2+} and Zn^{2+} were strong inhibitors which resulted in only 31% residual *Am*-fluorinase activity. In contrast, Mg^{2+} slightly increased fluorination activity (+ 9.5%) and EDTA slightly reduced activity (- 8.9%) in comparison to the control.

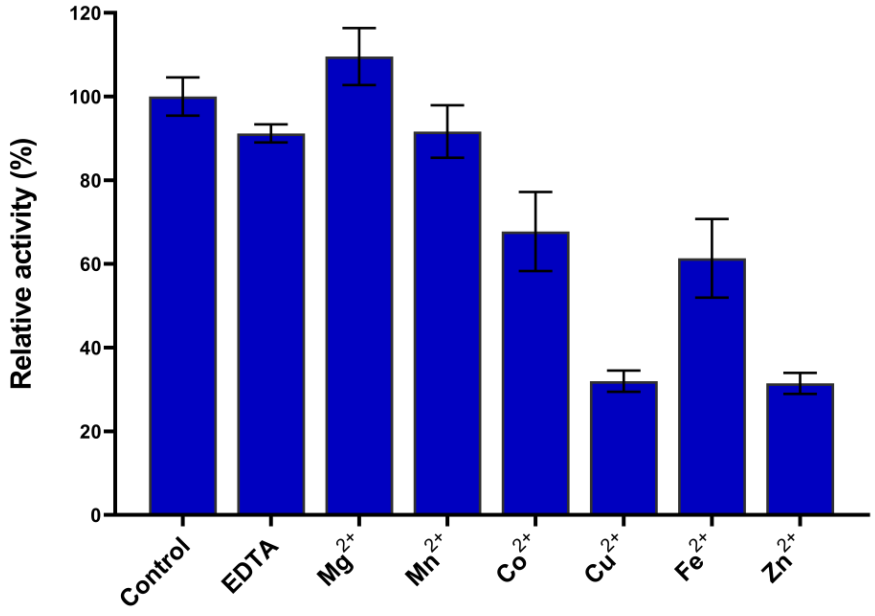


Fig. 7: The influence of divalent metal ions on fluorinase activity. Various metal ions or EDTA, a chelator, were used at a concentration of 1 mM. Treatments were compared to the control assay which did not contain any metal ions or EDTA. Error bars represent the standard deviation obtained from a minimum of 3 replicates.

3.5.3 Enzyme stability

Enzyme stability is a highly desirable attribute for industrial biocatalysts. Therefore, the stability of the fluorinase at 4 °C and 25 °C was tested (Fig. 8). After storage for 2 months

at 4 °C, the fluorinase retained 84% of activity. But more intriguingly, 26% activity was still detected after 2 months at 25 °C.

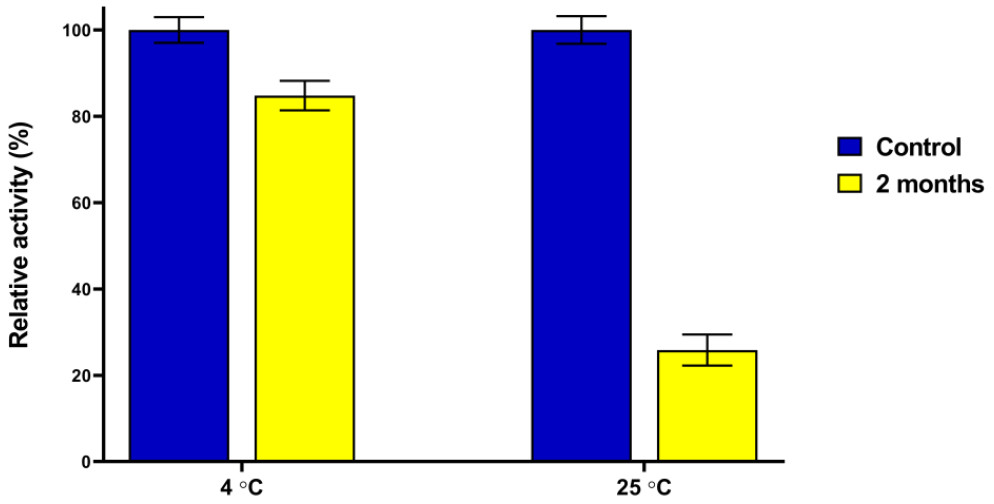


Fig. 8: Stability of the fluorinase at 4 ° and 25 °C. Activity of the fluorinase was measured directly after purification (day 0 or control). The enzyme was then maintained at either 4 °C or 25 °C for 2 months, after which, fluorination activity was tested again. Error bars represent the standard deviation obtained from 3 replicates.

4. Discussion

With the capacity to catalyse a C-F bond, fluorinases have the potential to be applied as biocatalysts in the preparation of fine chemicals [8]. But these enzymes are rare and, presently, only five representatives are available [6,19-21]. Sequence homology searches led to the identification of a putative fluorinase from *A. mzabensis*. Although annotated as a hypothetical protein, sequence alignments with other SAM-utilising enzymes suggested that the *Am*-fluorinase contained characteristic features (Fig. 2) of fluorinases. Similar to the known fluorinases, the *Am*-fluorinase contained a 22 amino acid motif which translates into a distinct loop in the three-dimensional monomeric structure (Fig. S3). The 22 amino acid loop has been proposed as a distinguishing feature of bacterial fluorinases [7,34]. Although the exact function of this loop is still unclear as it does not bind any substrate directly [8]; a

portion of the loop (A95-Q102) has been implicated in the formation of the binding site for the $_{L}$-methionine moiety of SAM [34]. In addition, residues that play a central role in catalysis were highly conserved in the *Am*-fluorinase. D16 is essential for the positioning and binding of SAM within the active site while F156 facilitates desolvation of the fluoride ion by providing the rear of the hydrophobic pocket. Mutations of these residues were found to be detrimental for fluorinase activity [21,34]. S158 is a key residue in halide binding and mutants at this position typically retain less than 38% activity. T80 together with Y77 have also been suggested to play a structural role in the halide binding site [21,34]. Collectively, the bioinformatics results provided compelling evidence to investigate the *Am*-fluorinase sequence as a potential fluorinase.

In order to confirm fluorination activity, the *Am*-fluorinase was cloned into *E. coli* cells. During overexpression, however, the *Am*-fluorinase accumulated as insoluble IBs (Fig. 3A). This is consistent with previous reports in which expression of fluorinases were also exceedingly prone to the formation of IBs [19,21]. Enzymes within IBs are usually devoid of catalytic activity and the attainable yield of active or soluble target protein is severely limited [22]. But despite this, all earlier studies have opted to purify the fluorinase from the soluble protein fraction [7,19-21,34] In this study, a method to recover and renature the *Am*-fluorinase from IBs was established instead.

For most proteins, aggregation during the refolding process is a key challenge. This accounts for majority of the loss in protein yield from IBs [35]. Aggregation has been attributed to the non-specific interaction of partially refolded molecules [35] and reports have shown that host cell contaminants such as nucleic acids, proteins and phospholipids influence this process [36-38]. Prior to IB denaturation and refolding, efforts were made to remove some impurities (Fig. 3B). The use of triton-X 100 allowed for cellular membrane debris and some membrane-associated proteins to be solubilised while IBs remained in the pellet following centrifugation [37]. An additional wash with urea was then performed. Urea, at moderate concentrations, has been shown to selectively solubilise some contaminating proteins that coaggregate with the target protein [37,39]. Following partial removal of contaminants, the *Am*-fluorinase was the major constituent (59%) in IBs. This, coupled with the gradual removal of denaturant (8 M urea) in the presence of a stabilising agent (glycerol), appeared to circumvent aggregation during the refolding process. Consequently, most of the *Am*-fluorinase was in a soluble form. Partial removal of contaminants also reduced the number of purification steps needed to obtain pure protein.

Following a single purification step (Fig. 4A), we achieved improved quantities of the fluorinase at a purity suitable for biocatalytic applications. An additional gel filtration step resulted ≥ 95% pure *Am*-fluorinase (Fig. 4B).

Activity assays validated the identification of a new fluorinase from *A. mzabensis* (Fig. 5). Moreover, these results indicated that the *Am*-fluorinase was refolded into its native form, or that it at least assumed a conformation capable of binding and catalysing SAM into 5'-FDA. In solution, previous fluorinases were found to exist as hexamers [6,20,21]. They consist of six monomeric subunits (~ 32 kDa each) that natively fold into a dimer of trimers and substrate binding occurs at the interface of two monomers [7,34]. The native size can provide insights into the structure that the refolded *Am*-fluorinase assumes in solution but this is yet to be determined. Nevertheless, the refolded *Am*-fluorinase was biologically active. And the method presented here for the recovery of the fluorinase from IBs can potentially be applied to other fluorinases.

Characterisation of the *Am*-fluorinase was conducted to explore the potential range of application for the enzyme. Preliminary results suggest the optimal pH for the *Am*-fluorinase is about 7.2 whereas the optimum temperature is around 65 °C (Fig. 6). However, additional points are recommended in order to pinpoint the exact values. In contrast, more acidic optima were observed for the fluorinases from *S. cattleya*, *N. brasiliensis* and *S. xinghaiensis* at pH 7.0, pH 6.5 and pH 5.0, respectively [6,20,21]. Furthermore, optimal temperatures were observed at 55 °C for *S. cattleya* [6] and 40 °C for *S. xinghaiensis* [21]. Mg^{2+} was found to slightly increase the activity of the *Am*-fluorinase (Fig. 7). In the absence of any metal ions though, activity of the *Am*-fluorinase was only slightly reduced (- 8.9%) and not completely abolished. Ma et al observed similar findings for the *S. xinghaiensis* fluorinase, whereby EDTA reduced activity by almost 40%, and hence concluded that their fluorinase may be metal ion dependent [21]. However, given that activity was not abolished entirely, our findings suggest that activity of the *Am*-fluorinase is not dependent on metal co-factors. Metals ions are not directly involved in catalysis but the exact interaction with fluorinases is still unclear. We propose that some ions induce conformational changes in the native structure of the enzyme instead, which results in slightly more efficient activities. With respect to enzyme stability, a 74% loss in *Am*-fluorinase activity was seen after 2 months at 25 °C (Fig. 8), whereas a 13% loss in activity after 4 days at 25 °C was reported for the *S. xinghaiensis* fluorinase [21]. Taken together with the optimal temperature experiments, the data presented here demonstrates a conceivably thermostable fluorinase

from *A. mzabensis*. It is not unusual for enzymes from organisms isolated in environments with elevated temperatures to display improved thermal tolerance [40], as it may be the case for the fluorinase from *A. mzabensis* [23]. Accordingly, the *Am*-fluorinase is a promising candidate for biocatalytic applications.

Conclusion

Our findings confirm the identification of an additional fluorinase from *Actinopolyspora mzabensis*. Expression of the recombinant enzyme in *E. coli* resulted in the formation of inclusion bodies. Subsequent washing, solubilisation and refolding of IBs allowed for the recovery of the *Am*-fluorinase in a biologically active form. Furthermore, improved yields (> 80 mg) were obtained following a single purification step using hydrophobic interaction chromatography. To the best of our knowledge, this is the first report on the successful refolding of a fluorinase. The method presented here can potentially be applied to other fluorinases for improved recovery yields. The *Am*-fluorinase identified here displayed a unique set of conditions for optimal activity (pH 7.2 and temperature of 65 °C) and exhibited excellent thermostability. As such, this enzyme is a promising candidate for biocatalytic applications. The identification of a new fluorinase contributes to the resources available for biofluorination and facilitates the development of sustainable catalysis.

Acknowledgements

We thank the Mass Spectrometry Facility (School of Chemistry, University of the Witwatersrand) for use of the LC-MS, as well as Eric Morifi and Thapelo Mbhele for training and assistance with the instrument. We also kindly thank Dr. Allan Prior (now at the University of Hawaii) for the synthetic 5'-FDA sample. This work is based on the research supported in part by the National Research Foundation of South Africa (Grant No. 111329 to S.A.S). The opinions, findings and conclusions or recommendations expressed in this publication are that of the author(s) and the NRF accepts no liability whatsoever in this regard.

Competing interests

The authors declare no competing interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Wang, J., Sánchez-Roselló, M., Aceña, J. L., del Pozo, C., Sorochinsky, A. E., Fustero, S., Soloshonok, V.A. and Liu, H. (2013) Fluorine in pharmaceutical industry: fluorine-containing drugs introduced to the market in the last decade (2001–2011) *Chemical reviews*, *114*, 2432-2506.
2. Fujiwara, T. and O'Hagan, D. (2014) Successful fluorine-containing herbicide agrochemicals. *Journal of Fluorine Chemistry*, *167*, 16-29.
3. Furuya, T., Kamlet, A. S. and Ritter, T. (2011) Catalysis for fluorination and trifluoromethylation. *Nature*, *473*, 470.
4. Liang, T., Neumann, C. N. and Ritter, T. (2013) Introduction of Fluorine and Fluorine-Containing Functional Groups. *Angewandte Chemie International Edition*, *52*, 8214-8264.
5. O'Hagan, D., Schaffrath, C., Cobb, S. L., Hamilton, J. T. and Murphy, C. D. (2002) Biosynthesis of an organofluorine molecule. *Nature*, *416*, 279.
6. Schaffrath, C., Deng, H. and O'Hagan, D. (2003) Isolation and characterisation of 5′-fluorodeoxyadenosine synthase, a fluorination enzyme from *Streptomyces cattleya*. *FEBS letters*, *547*, 111-114.
7. Dong, C., Huang, F., Deng, H., Schaffrath, C., Spencer, J. B., O'Hagan, D. and Naismith, J. H. (2004) Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature*, *427*, 561.
8. O'Hagan, D. and Deng, H. (2014) Enzymatic fluorination and biotechnological developments of the fluorinase. *Chemical reviews*, *115*, 634-649.

9. Ma, L., Bartholome, A., Tong, M. H., Qin, Z., Yu, Y., Shepherd, T., Kyeremeh, K., Deng, H. and O'Hagan, D. (2015) Identification of a fluorometabolite from *Streptomyces* sp. MA37: (2*R*, 3*S*, 4*S*)-5-fluoro-2,3,4-trihydroxypentanoic acid. *Chemical science*, *6*, 1414-1419.
10. Truppo, M. D. (2017) Biocatalysis in the pharmaceutical industry: the need for speed. *ACS medicinal chemistry letters*, *8*, 476-480.
11. Martarello, L., Schaffrath, C., Deng, H., Gee, A. D., Lockhart, A. and O'Hagan, D. (2003) The first enzymatic method for C–^{18}F bond formation: the synthesis of 5′-[^{18}F]-fluoro-5′-deoxyadenosine for imaging with PET. *Journal of Labelled Compounds and Radiopharmaceuticals*, *46*, 1181-1189.
12. Deng, H., Cobb, S. L., Gee, A. D., Lockhart, A., Martarello, L., McGlinchey, R. P., O'Hagan, D. and Onega, M. (2006) Fluorinase mediated C–^{18} F bond formation, an enzymatic tool for PET labelling. *Chemical communications*, 6, 652-654.
13. Winkler, M., Domarkas, J., Schweiger, L. F. and O'Hagan, D. (2008) Fluorinase-coupled base swaps: synthesis of [^{18}F]-5′-deoxy-5′-fluorouridines. *Angewandte Chemie International Edition*, *47*, 10141-10143.
14. Thompson, S., Zhang, Q., Onega, M., McMahon, S., Fleming, I., Ashworth, S., Naismith, J.H., Passchier, J. and O'Hagan, D. (2014) A Localized Tolerance in the Substrate Specificity of the Fluorinase Enzyme enables "Last-Step" ^{18}F Fluorination of a RGD Peptide under Ambient Aqueous Conditions. *Angewandte Chemie International Edition*, *53*, 8913-8918.
15. Li, X. G., Domarkas, J. and O'Hagan, D. (2010) Fluorinase mediated chemoenzymatic synthesis of [^{18}F]-fluoroacetate. *Chemical Communications*, *46*(41), 7819-7821.
16. Onega, M., Domarkas, J., Deng, H., Schweiger, L. F., Smith, T. A., Welch, A. E., Plisson, C., Gee, A.D. and O'Hagan, D. (2010) An enzymatic route to 5-deoxy-5-[^{18}F] fluoro-d-ribose, a [^{18} F]-fluorinated sugar for PET imaging. *Chemical Communications*, *46*, 139-141.
17. Eustáquio, A. S., O'Hagan, D. and Moore, B. S. (2010) Engineering fluorometabolite production: fluorinase expression in *Salinispora tropica* yields fluorosalinosporamide. *Journal of Natural Products*, *73*, 378-382.
18. Walker, M. C., Thuronyi, B. W., Charkoudian, L. K., Lowry, B., Khosla, C. and Chang, M. C. (2013) Expanding the fluorine chemistry of living systems using engineered polyketide synthase pathways. *Science*, *341*, 1089-1094.

19. Deng, H., Ma, L., Bandaranayaka, N., Qin, Z., Mann, G., Kyeremeh, K., Yu, Y., Shepherd, T. and O'Hagan, D. (2014) Identification of fluorinases from *Streptomyces* sp MA37, *Norcardia brasiliensis*, and *Actinoplanes* sp N902-109 by Genome Mining. *ChemBioChem*, *15*, 364-368.
20. Wang, Y., Deng, Z. and Qu, X. (2014) Characterization of a SAM-dependent fluorinase from a latent biosynthetic pathway for fluoroacetate and 4-fluorothreonine formation in *Nocardia brasiliensis*. *F1000Research*, *3*.
21. Ma, L., Li, Y., Meng, L., Deng, H., Li, Y., Zhang, Q. and Diao, A. (2016) Biological fluorination from the sea: discovery of a SAM-dependent nucleophilic fluorinating enzyme from the marine-derived bacterium *Streptomyces xinghaiensis* NRRL B24674. *RSC Advances*, *6*, 27047-27051.
22. Singh, S. M. and Panda, A. K. (2005) Solubilization and refolding of bacterial inclusion body proteins. *Journal of bioscience and bioengineering*, *99*, 303-310.
23. Meklat, A., Bouras, N., Zitouni, A., Mathieu, F., Lebrihi, A., Schumann, P., Spröer, C., Klenk, H.P. and Sabaou, N. (2013) *Actinopolyspora mzabensis* sp. nov., a halophilic actinomycete isolated from an Algerian Saharan soil. *International Journal of Systematic and Evolutionary Microbiology*, *63*, 3787-3792.
24. 2ndFind (2018) A web based support tool to find secondary metabolic biosynthetic gene clusters. http://biosyn.nih.go.jp/2ndfind/
25. Webb, B. and Sali, A. (2014) Comparative protein structure modeling using MODELLER. *Current protocols in bioinformatics*, *47*, 5-6
26. DeLano, W. L. (2002) Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*, *40*, 82-92.
27. Inoue, H., Nojima, H. and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, *96*, 23-28.
28. Schägger, H. (2006) Tricine–SDS-PAGE. *Nature Protocols*, *1*, 16.
29. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. and Tinevez, J. Y. (2012) Fiji: an open-source platform for biological-image analysis. *Nature Methods*, *9*, 676.
30. Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*, 571-607, Humana press.

31. Eustáquio, A. S., Pojer, F., Noel, J. P. and Moore, B. S. (2008) Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nature chemical biology*, *4*, 69.
32. Deng, H., Botting, C. H., Hamilton, J. T., Russell, R. J. and O'Hagan, D. (2008) *S*-Adenosyl-$_{L}$ -methionine: Hydroxide Adenosyltransferase: A SAM enzyme. *Angewandte Chemie*, *120*, 5437-5441.
33. Deng, H. and O'Hagan, D. (2008) The fluorinase, the chlorinase and the duf-62 enzymes. *Current Opinion in Chemical Biology*, *12*, 582-592.
34. Zhu, X., Robinson, D. A., McEwan, A. R., O'Hagan, D. and Naismith, J. H. (2007) Mechanism of Enzymatic Fluorination in *Streptomyces cattleya*. *Journal of the American Chemical Society*, *129*, 14597-14604.
35. Tsumoto, K., Ejima, D., Kumagai, I. and Arakawa, T. (2003) Practical considerations in refolding proteins from inclusion bodies. *Protein Expression and Purification*, *28*, 1-8.
36. Maachupalli-Reddy, J., Kelley, B. D. and Clark, E. D. B. (1997) Effect of inclusion body contaminants on the oxidative renaturation of hen egg white lysozyme. *Biotechnology Progress*, *13*, 144-150.
37. Batas, B., Schiraldi, C. and Chaudhuri, J. B. (1999) Inclusion body purification and protein refolding using microfiltration and size exclusion chromatography. *Journal of Biotechnology*, *68*, 149-158.
38. Fahey, E. M., Chaudhuri, J. B. and Binding, P. (2000) Refolding and purification of a urokinase plasminogen activator fragment by chromatography. *Journal of Chromatography B: Biomedical Sciences and Applications*, *737*, 225-235.
39. Shi, T., Zhang, L., Li, Z., Newton, I. P. and Zhang, Q. (2015) Expression, purification and renaturation of truncated human integrin β1 from inclusion bodies of *Escherichia coli*. *Protein Expression and Purification*, *107*, 13-19.
40. Haki, G. D. and Rakshit, S. K. (2003) Developments in industrially important thermostable enzymes: a review. *Bioresource Technology*, *89*, 17-34.

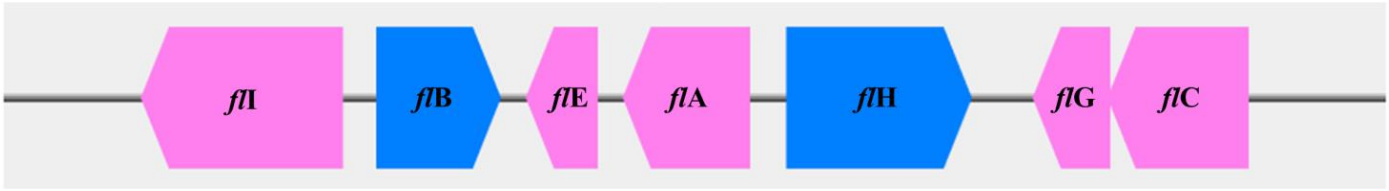
Identification and characterisation of a novel fluorinase from *Actinopolyspora mzabensis*

Selisha Ann Sooklal[a] and Karl Rumbold[a*]

[a] Department of Microbiology, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, 2000, South Africa

[*] Corresponding author: Department of Microbiology, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, 2000, South Africa. E-mail: karl.rumbold@wits.ac.za

Supplementary information



Gene	Length (aa)	Proposed gene product
*fl*I	484	*S*-adenosyl-$_L$-homocysteine hydrolase
*fl*B	294	*S*-methyl-5'-thioadenosine phosphorylase
*fl*E	166	Adenine phosphoribosyl transferase
*fl*A	299	5'-fluoro-5'-deoxyadenosine synthase (*Am*-fluorinase used in this study)
*fl*H	442	Na^+/ H^+ antiporter
*fl*G	178	DNA binding protein (regulatory)
*fl*C	333	Permease

Fig. S1: An 8kb fragment of the fluorometabolite gene cluster found in *A. mzabensis* (GenBank accession: FNFM01000017.1) using the 2ndFind tool. The cluster is homologous to the fluorometabolite gene clusters found in *S. cattleya*, *Streptomyces* sp MA37, *Actinoplanes* sp, *N. brasiliensis* and *S. xinghaiensis*. The cluster shows the arrangement of seven genes (blue on the sense (+) strand and pink on the antisense (-) strand) while the corresponding table provides the associated length and proposed product for each gene.

Table S1: GenBank accession numbers and sequences for all enzymes used in this study.

Accession number	Sequence
Fluorinases	
SDK90789.1 Description: hypothetical protein from *Actinopolyspora mzabensis*	MSDSYSRPIIAFMSDLGTTDDSVAQCKGLMMSICQDVTVVDVCHSMEPWN VEEGARYIVDLPRFFPEGTVFATTTYPATGTTARSVAVRIKYPAKGGARG QWAGSGEGFERSEGSYIYIAPNNGLLTTVLQEHGYTEAYEVSSTDVVPAR PEPTFYSREMVAIPSAHLAAGYPLEKVGRKLQDSEIVRFTPPQATVSPEG DLSGVVTAIDHPFGNIWTSIHRDNLESAGVGYGTNLKIVLDDVFPFELPL SPTFADAGEVGDPVVYVNSRGYLSLARNAASLAYPYNLKEGMSVRVTRS
WP_014144878.1 Description: adenosyl-fluoride synthase from *Streptomyces cattleya*	MAANSTRRPIIAFMSDLGTTDDSVAQCKGLMYSICPDVTVVDVCHSMTPW DVEEGARYIVDLPRFFPEGTVFATTTYPATGTTTRSVAVRIKQAAKGGAR GQWAGSGAGFERAEGSYIYIAPNNGLLTTVLEEHGYLEAYEVTSPKVIPE QPEPTFYSREMVAIPSAHLAAGFPLSEVGRPLEDHEIVRFNRPAVEQDGE ALVGVVSAIDHPFGNVWTNIHRTDLEKAGIGYGARLRLTLDGVLPFEAPL TPTFADAGEIGNIAIYLNSRGYLSIARNAASLAYPYHLKEGMSARVEAR
CDH39444.1 Description: fluorinase from *Streptomyces* sp. MA37	MAANGSQRPIIAFMSDLGTTDDSVAQCKGLMHSICPGVTVVDVCHSMTPW DVEEGARYIVDLPRFFPEGTVFATTTYPATGTTTRSVAVRIRQAAKGGAR GQWAGSGDGFERADGSYIYIAPNNGLLTTVLEEHGYIEAYEVTSTKVIPA NPEPTFYSREMVAIPSAHLAAGFPLAEVGRRLDDSEIVRFHRPAVEISGE ALSGVVTAIDHPFGNIWTNIHRTDLEKAGIGQGKHLKIILDDVLPFEAPL TPTFADAGAIGNIAFYLNSRGYLSLARNAASLAYPYNLKAGLKVRVEAR
WP_019711456.1 Description: adenosyl-fluoride synthase from *Streptomyces xinghaiensis*	MSADPTQRPIIGFMSDLGTTDDSVAQCKGLMHSICPGVTVIDVCHSMTPW DVEEGARYIVDLPRFFPEGTVFATTTYPATGTETRSVAVRIKQAAKGGAR GQWAGSAGGFERAEGSYIYVAPNNGLLTTVLEEHGYIEAYEVSSTKVIPE RPEPTFYSREMVAIPAAHLAAGFPLSEVGRPLEDSEIVRYQPPQVEISGD TLTGVVSAIDHPFGNVWTNIHRTHLEKAGIGYGKRIKIILDDVLPFEQTL VPTFADAGEIGGVAAYLNSRGYLSLARNAASLAYPFNLKAGLKVRVETN
WP_042263697.1 Description: adenosyl-fluoride synthase from *Nocardia brasiliensis*	MSDLGITDDSVAQCKGLMLSVCPDVTIVDICHTMQPWDVEEGARYIVDLP RLFPEGTVFATTTYPATGTTARSVALRIAHASKGGARGQWAGSGAGFERK EGSYIYIAPNNGLLTTVIKEHGYLEAYEVSSPEVIPEQPEPTFYSREMVA LPSAHLAAGFPLEKVGRRLADDEIVRFERKDPELVADTDLVGYVTNIDHP FGNVWTNIHRTDLEKLGVGYGTKLRITLDGVLPFELPLSPTFADAGEVGA AVAYLSSRGYLALARNAASLAYPYNLKAGISVQVKVG
WP_041832040.1 Description: adenosyl-fluoride synthase from *Actinoplanes* sp. N902-109	MPANGNPIIAFMSDLGTTDDSVAQCKGLMLSICPGVTIVDVNHSMTPWDV EEGARYIVDLPRFFPEGTVFATTTYPATGTATRSVALRIKQAAQGGARGQ WAGSGAGFERAEGSYIYIAPNNGLLTTVIEEHGYIEAYEVSNTKVIPAEP EPTFYSREMVAIPSAHLAAGFPLNEVGRALSDDEIVRFAKPKPSTVSGGV LSGVITNIDHPFGNLWTNIHRTDLEKAGIGYQTQLRLLLDGVLTFDLPLV PTFADAGQIGDPVIYINSRGYLALARNAAPLAYPYNLKAGLTVTVTKA
Chlorinase	
ABP73643.1 Description: SalL from *Salinispora tropica*	MQHNLIAFLSDVGSADEAHALCKGVMYGVAPAATIVDITHDVAPFDVREG ALFLADVPHSFPAHTVICAYVYPETGTATHTIAVRNEKGQLLVGPNNGLL SFALDASPAVECHEVLSPDVMNQPVTPTWYGKDIVAACAAHLAAGTDLAA VGPRIDPKQIVRLPYASASEVEGGIRGEVVRIDRAFGNVWTNIPTHLIGS MLQDGERLEVKIEALSDTVLELPFCKTFGEVDEGQPLLYLNSRGRLALGL NQSNFIEKWPVVPGDSITVSPRVPDSNLGPVLG

Table S1: continued

Hydroxide adenosyl transferase	
PDB: 2WR8 Description: SAM hydroxide-adenosyltransferase from *Pyrococcus horikoshii*	GSHMITLTTDFGLKGPYVGEMKVAMLRINPNAKIVDVTHSVTRHSILEGS FVMEQVVKYSPKGTVHVGVIDPGVGTERRAIVIEGDQYLVVPDNGLATLP LKHIKVKSVYEIIPDKIRKFTGWEISSTFHGRDIFGPAGALIEKGIHPEE FGREIPVDSIVKLNVEPRKEGDVWILKVIYIDDFGNVILNLENYEKPRTV ELLDFNLRLPYLETYGLVEKGEMLALPGSHDYLEIAVNMGSAAERLNVKV GDELRVRLL
WP_088854816.1 Description: hydroxide adenosyltransferase from *Thermococcus pacificus*	MITLTTDFGLKGPYVGEMKVAMLRVNPNAKLVDVTHAVTRHSIVEGSFVM EQVVKYSPEGTVHVGVIDPGVGTERRAVIVEGEQWLVVPDNGLATLPMKH IKPRRAWEIDLERIRRFTGWKISSTFHGRDVFGPAGALIEKGVSPEEFAE EVPLDSLVRLDLKPRKEKDLWLLKVIYVDDFGNVILNLEDYERLNAVELP DFRLRIPYLDTYGQVKPGELLALPGSHDYLEIAVNQGSASERLGLKVGDE VRVRFVQKF
WP_048148689.1 Description: hydroxide adenosyltransferase from *Palaeococcus ferrophilus*	MITLTTDFGLRGPYVGEMKVAMLRVNPKATIIDVTHGITRHSIIEGSFVM EQVVKYSPEGAVHVGVIDPGVGTSRRAVVIEGNQWLVVPDNGLATLPLKH IKPKRAWEIDVEKLKRFTGWEVSSTFHGRDVFGPAGALIDREISPEEFAT EIPLESLIKLDVEPRKEGDTWILKVIYVDDFGNVILNLEDYGRPSEVELP DFNLKIPYLETYGMVNRGELLALPGSHGYLEIAVNQGSAAGRLGLNVGDE VRVRLI

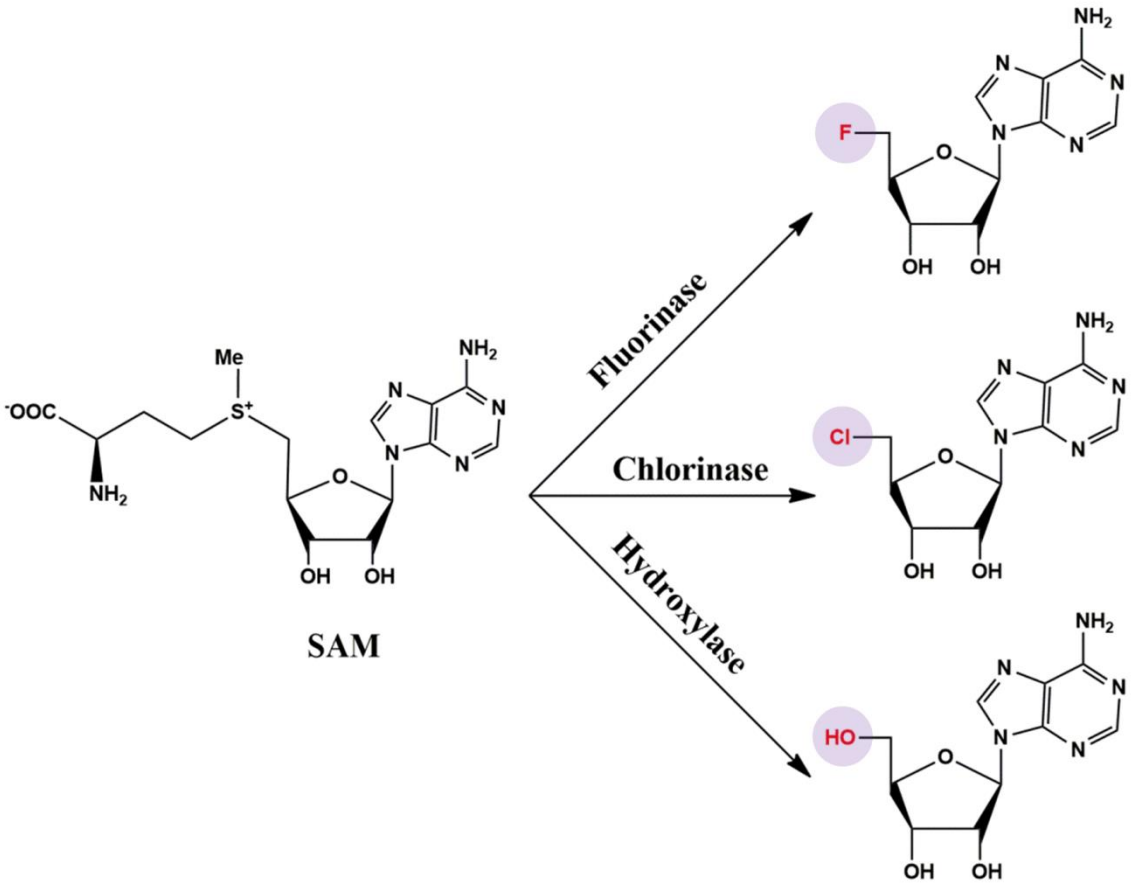


Fig. S2: Comparison of the reactions catalysed by fluorinases, chlorinases and hydroxide adenosyltransferases. Fluorinases selectively catalyse the incorporation of fluoride ions (F^-) at the C-5 position of SAM while chlorinases incorporate chloride (Cl^-) and hydroxide adenosyltransferases incorporate a hydroxyl group (OH^-) at the same position. Structures produced using ChemDraw version 12.0.2.

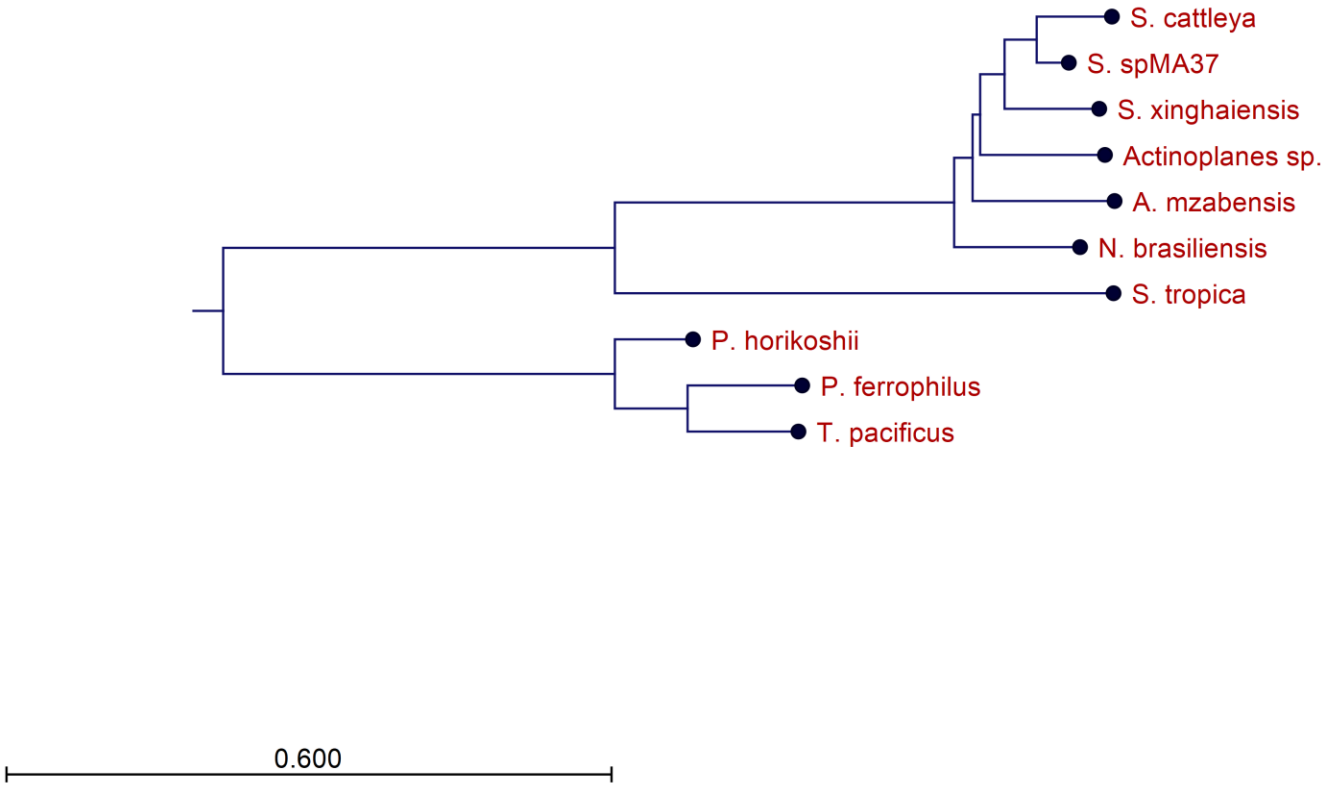


Fig. S3: Phylogenetic relationship between fluorinases (*S. cattleya*, *Streptomyces* sp. MA37, *Streptomyces xinghaiensis*, *Actinoplanes* sp. and *N. brasiliensis*), **the chlorinase** (*S. tropica*) **and hydroxide adenosyltransferases** (*P. horikoshii*, *P. ferrophilus* and *T. pacificus*). The putative fluorinase from *A. mzabensis* (used in this study) forms a distinct group with the five known fluorinases. The tree was created using the CLC Genomics Workbench version 12.0 (Qiagen) using the neighbor-joining method.

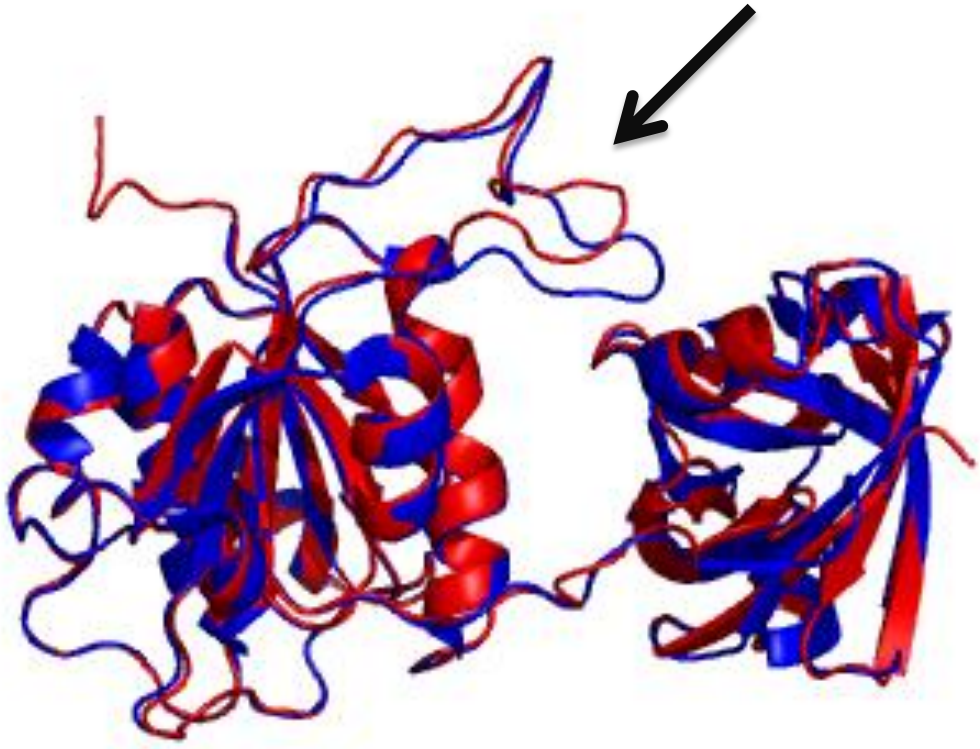


Fig. S4: Structure of the *A. mzabensis* fluorinase monomer (red) modelled against the fluorinase from *S. cattleya* (blue). The structure was created *in silico* using MODELLER version 9.19 by employing the fluorinase from *S. cattleya* (PDB accession: 1RQP) as a template. The 22 amino acid motif, characteristic of fluorinases, forms a loop structure and is indicated by the black arrow.

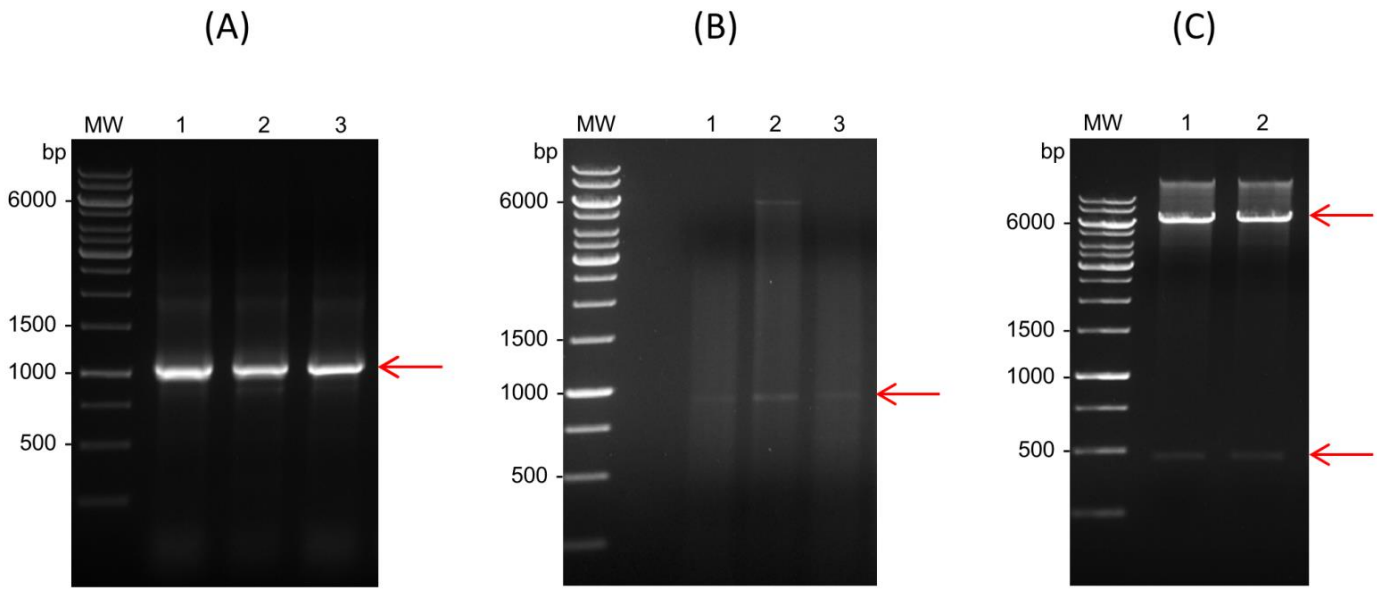


Fig. S5: Colony PCR and restriction digests used to screen for positive transformants. **(A)** Colony PCR using universal T7 primers (F: TAATACGACTCACTATAGGG; R: GCTAGTTATTGCTCAGCGG) to screen for *E. coli* cells harbouring the pET11a-*Am*-fluorinase plasmid. PCR products were resolved on a 1% agarose gel. This resulted in a band of ~ 1kb (red arrow), consistent with the expected size. **(B)** Plasmid DNA was extracted from transformants (shown in A) using the Zymo ZR Plasmid miniprep kit and restriction digests (*Nde*I and *Bam*HI) were performed to further confirm the pET-11a-*Am*-fluorinase plasmid. Restriction products were resolved on a 1% agarose gel. Bands corresponding to the *Am*-fluorinase insert (expected size ~ 900bp) and pET11a plasmid (expected size ~ 5.7 kb) were seen. **(C)** Restrictions digests were also performed using *Kpn*I. This enzyme was expected to cut once within the *Am*-fluorinase gene and once on the pET11a vector resulting in fragments of ~6 kb and ~ 500bp, as indicated by the red arrows.

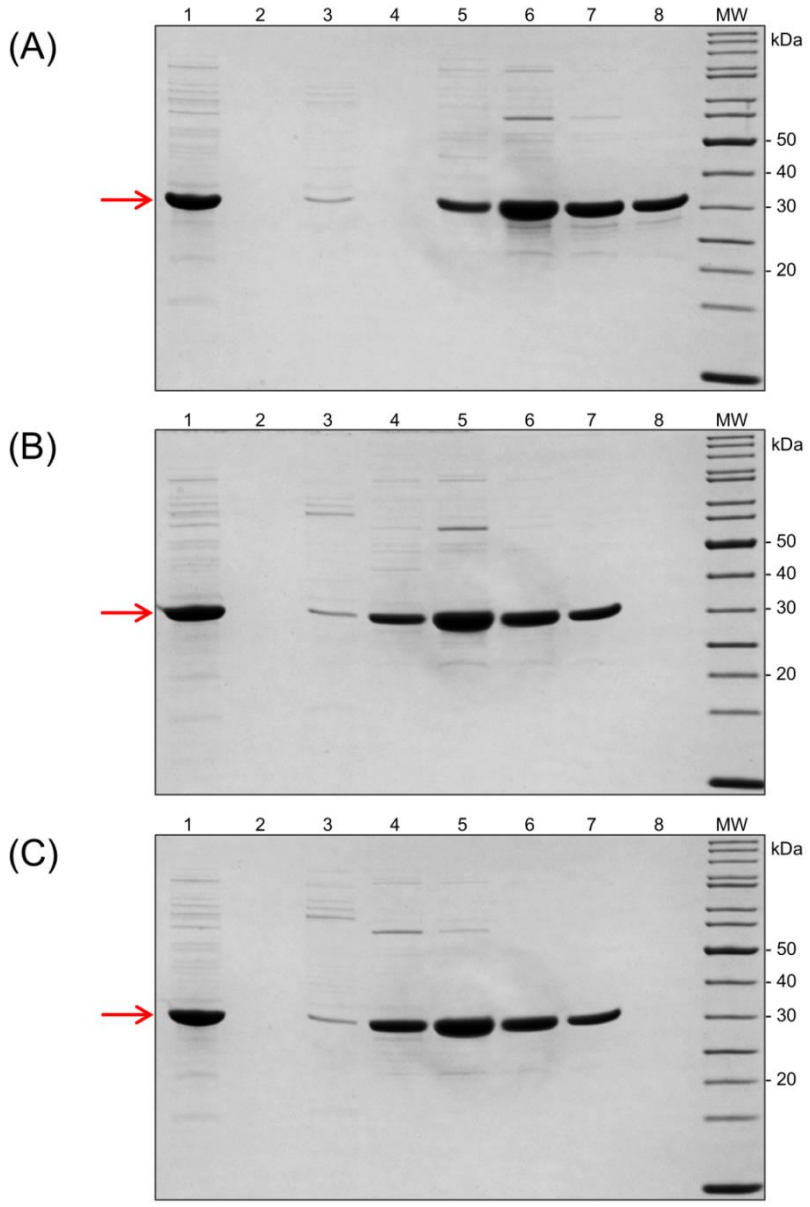


Fig. S6: Optimisation of NaCl concentrations (400 mM, 200 mM and 100 mM) for hydrophobic interaction chromatography. (A) Refolded *Am*-fluorinase samples were pre-equilibrated with 400 mM NaCl and loaded (lane 1) onto a phenyl sepharose column. Bound protein was eluted in a single step with 0 mM NaCl in the buffer. Fractions were collected (lanes 3, 5-8) and resolved on a 14% tricine-SDS-PAGE gel. **(B)** The refolded *Am*-fluorinase was pre-equilibrated with 200 mM NaCl and loaded (lane 1) onto a phenyl sepharose column as before. Fractions (lanes 3-7) were collected and resolved on a gel. **(C)** The refolded *Am*-fluorinase was loaded (lane 1) directly onto the phenyl sepharose column, since 100 mM NaCl was included in the refolding buffer, and fractions (lanes 3-7) were collected. The *Am*-fluorinase (indicated by the red arrow) was able to bind to the column at all NaCl concentrations tested, however, more selective binding of the target protein (in comparison to contaminants) was achieved with 100 mM NaCl.

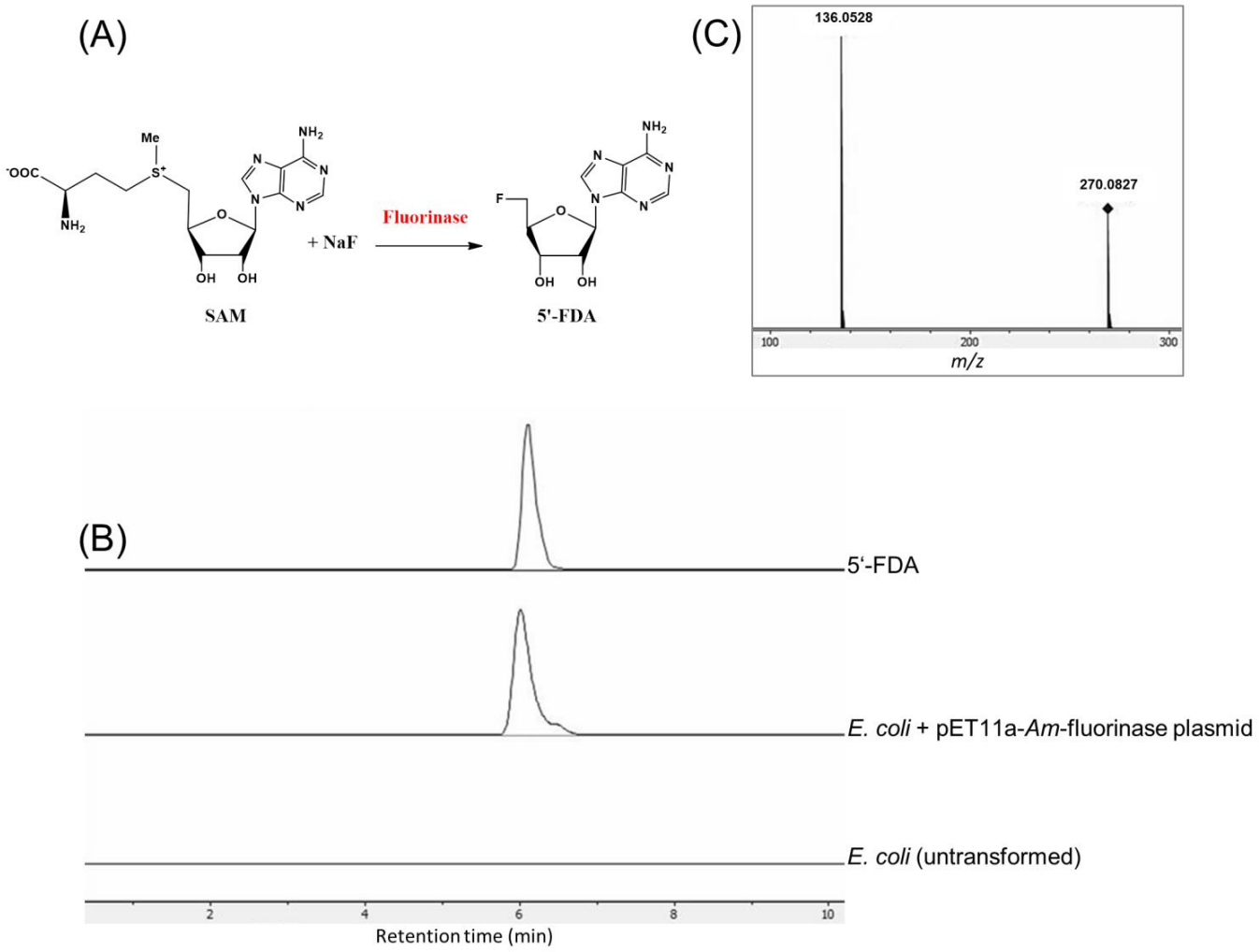


Fig. S7: Chromatogram of fluorination activity for *E. coli* (transformed and untransformed) crude cell lysates. (A) The reaction catalysed exclusively by fluorinases. **(B)** LC-MS/MS analysis of synthetic 5'-FDA (positive control) as well as the reaction products of the crude cell lysate of either *E. coli* cells transformed with the pET11a-*Am*-fluorinase plasmid or untransformed *E. coli* cells. The untransformed cells were unable to produce 5'-FDA. **(C)** The corresponding mass spectrum for the peak produced in (B) showing the MRM transition of 5'-FDA (*m/z* 270.0) into the fragment ion (*m/z* 136.0).

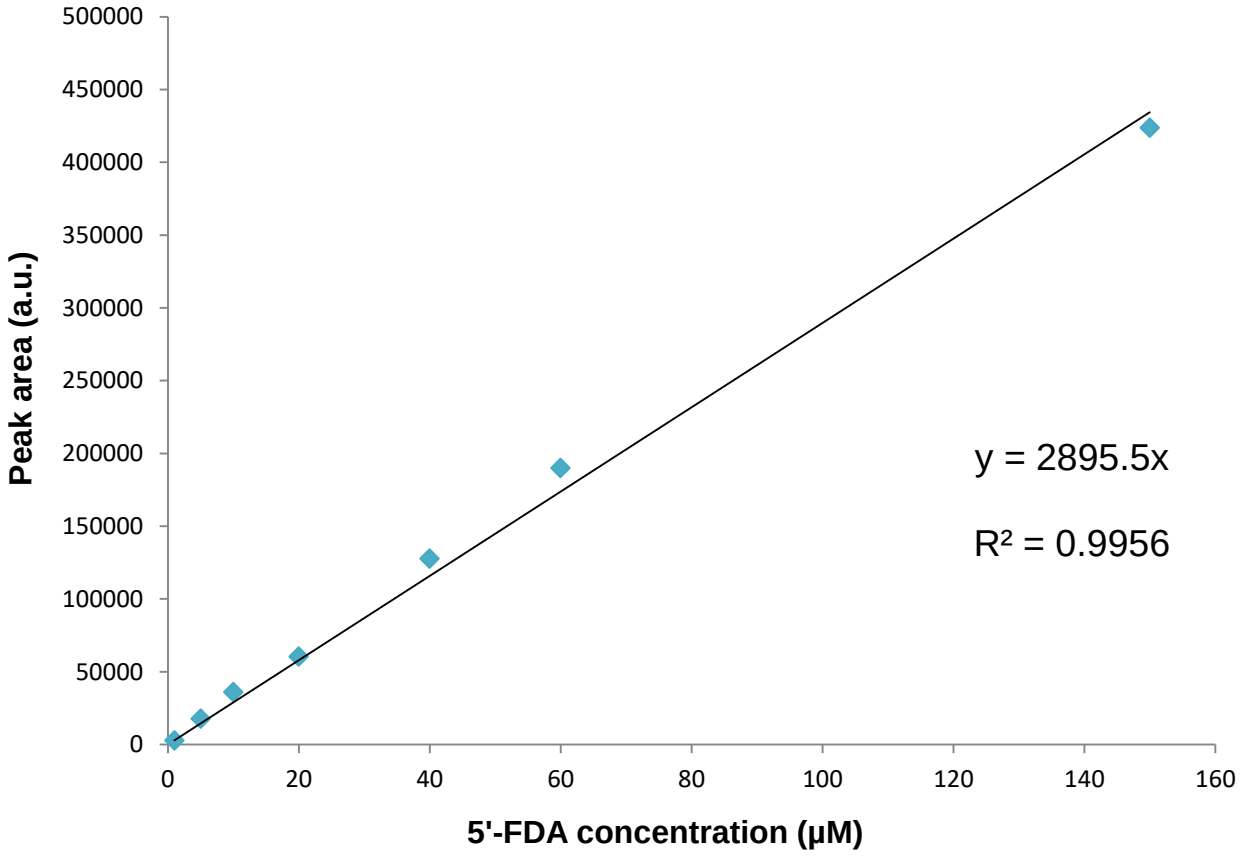


Fig. S8: Standard curve of synthetic 5'-FDA at various concentrations (1 µM, 5 µM, 10 µM, 20 µM, 40 µM, 60 µM and 150 µM) using an optimised LC-MS/MS method.

Chapter 5

General discussion and conclusions

5.1 The search for fluorinases in plants

It has been 76 years since the isolation of fluoroacetate from *D. cymosum* [50]. In this time, the production of FA has been confirmed in almost 60 other plants [80]. However, the mechanism in which C-F bonds are formed in plants has still not been elucidated. Furthermore, previous research into *D. cymosum* is obsolete, predating the discovery of the first fluorinase from the bacterium *Streptomyces cattleya*. Recent advances in high-throughput technologies, such as RNA-seq and LC-MS, have allowed for the acquisition of unprecedented amounts of molecular data [127]. This work aimed at investigating *D. cymosum* on an omic level to potentially gain insights into the fluorination mechanism.

Unlike some eukaryotic organisms such as nematodes and algae, higher plants are known to have large genomes [128]. Sequencing of plant genomes is often unfeasible due to the sheer complexity of the genome resulting from non-coding regions and high sequence repeat regions. Moreover, a substantial amount of time, money and computing resources are required to assemble meaningful genomes [129]. However, it is normally unnecessary to know all of the sequence information in a plant, especially when the goal of a study is limited to identification of particular genes of interest. This makes RNA-sequencing an attractive alternative and valid choice for plants, since the focus is restricted to coding regions only, rather than the entire genome [127]. RNA-seq is a powerful tool in molecular biology allowing for genetic profiling of non-model organisms through the simultaneous detection of transcript abundance and diversity [130]. Transcriptomes provide the essential link between the genome and proteome, and previous studies have successfully used this tool for the identification of novel transcripts as well as for the elucidation of pathways [131,132].

Since the discovery of the fluorinase from *S. cattleya*, new information has become available such as the gene sequence and natural substrate of the fluorinase [97]. These aspects have not been investigated in *D. cymosum* prior to this work. In this study (Chapter 2), a sequence similarity-based approach was chosen to inspect for potential fluorinases in *D. cymosum*. The production of FA is widely believed to be a defence mechanism in plants [74], however, very little is actually known about this. Therefore, transcriptomes from the unwounded and wounded *D. cymosum* leaf tissue were sequenced. *D. cymosum* had essentially no sequence information publically available, thus a reference transcriptome was created and, overall, 77,845 transcripts were assembled. The transcriptome was extensively characterised resulting in a 69% annotation rate. Despite this, a transcript associated with fluorination activity was not found. The fluorinase from *S. cattleya* was then searched against the *D. cymosum* transcriptome to identify transcripts with homology that were not annotated or annotated fallaciously. However, this did not yield any significant results. Zhu et al also reported that there had been no sequence match between bacterial fluorinases and genomes from the plant kingdom, and suggested the evolution of more than one biological mechanism for fluoride incorporation [104]. Although, it is possible that the sequencing depth of the *D. cymosum* transcriptome may have been insufficient to detect transcripts with low abundance, the failure to detect any fluorinases may also support the hypothesis of a distinctly different fluorinating enzyme in plants.

RNA-seq technologies have far surpassed the correlating bioinformatic tools needed for the interpretation of the data. A limitation in the analysis is that annotation is based on sequence similarity and is primarily database-driven. This means that for a transcript to be accurately annotated, it must be similar to another sequence and it must be present on a public database. This limits identification to only currently known genes [127]. An issue arises when novel enzymes, with a completely different sequence to anything known, needs to be detected. The sequence may be present in the sequence-data but not annotated, resulting in potentially unique enzymes being overlooked. This would especially be true if plants utilised a different fluorination mechanism to bacteria.

Considering the aforementioned conundrum, sequence similarity-based approaches would fail for the identification of a new fluorinating enzyme in plants. Therefore, a functional screening approach coupled with proteomic sequencing was employed (Chapter 3).

Proteomes provide the essential link between the transcriptome and metabolome. Shotgun proteomics involves the proteolytic digestion of proteins into smaller peptides. Subsequent identification of peptides is inferred by linking the mass-to-charge ratio to proteins on public databases. Yet again, if a specific protein is not represented on a database, it evades detection. In this instance, species-specific information such as the transcriptome can be used as a custom database to enhance identification of proteins [127]. Proteomes from eukaryotes, however, are complex and display a high dynamic range [133]. Therefore, coupled with functional screening and partial purification, novel proteins can be identified.

In the present study (Chapter 3), proteomes of the leaf, flower and seed tissue of *D. cymosum* were produced. Overall, a total of 2,430 proteins were detected. But, we were unable to detect any fluorination activity impeding tracing and detection of the enzyme through functional screening. The failure to detect fluorinase activity could be the result of many reasons such as enzyme concentration, assay conditions and consumption of the reaction product by other enzymes. However, a fundamental limitation in this study has been that activity assays were based on the assumption that potential fluorinases from plants utilise the same substrate (SAM) as bacterial fluorinases. Since this is the first study attempting activity assays using *D. cymosum* extracts with SAM, we are unable to verify these observations; however, the inability to detect a transcript with homology to any known fluorinase coupled with the inability to detect fluorinase activity using SAM, may provide evidence for a unique fluorinating enzyme in *D. cymosum*.

5.2 Fluorinases from bacteria

Fluorinases from bacteria remain the only representatives of enzymes with the ability to catalyse a C-F bond. However, these enzymes are exceedingly rare with only five identified thus far [99,108,109,112]. Given the industrial significance of these enzymes, it is no surprise that fluorinases are highly sought after and that every new addition is notable [19].

A novel fluorinase was identified from *Actinopolyspora mzabensis* (Chapter 4). Heterologous expression in *E. coli* resulted in the formation of inclusion bodies. This is not uncommon as prior fluorinases have also been prone to inclusion bodies [108,112]. However, previous studies have traditionally opted to salvage the soluble protein using

affinity chromatography, despite the low yields. Using this approach, Wang et al reported a yield of 7.9 mg/ L and 9.2 mg/L for the fluorinases from *N. brasiliensis* and *S. cattleya*, respectively [109]. And Ma et al reported a yield of ~100 mg per gram of dried *E. coli* cells which was acquired from a 2 L culture (i.e. ~50 mg/ L) [112]. For biocatalytic applications, practical expression systems combined with suitable quantities of the target enzyme is a prerequisite. This motivated the development of an alternative strategy. A method for the recovery and renaturation of the fluorinase from inclusion bodies was presented in the chapter. Using the protein refolding method proposed here, we were able to obtain a yield of ~ 80 mg/ L at a purity suitable for biocatalytic application. This is an improvement to yields previously attained. Moreover, the enzyme was functionally active. All three components of SAM (adenosine, ribose and methionine moieties) are recognised when binding occurs at the interface of two monomeric units of the fluorinase. Dong et al reported that the quaternary structure of the fluorinase is therefore crucial for substrate recognition and catalysis [45]. Although we were unable to determine the native size of the refolded fluorinase from *A. mzabensis*, the enzyme evidently assumes a biologically active conformation in solution.

Preliminary characterisation of the *A. mzabensis* fluorinase indicated an optimal pH and temperature around 7.2 and 65 °C. The error bars observed can be attributed to variations in protein batches caused during refolding. The presence of Mg^{2+} ions was found to increase relative activity by 10%. In contrast, the absence of Mg^{2+} did not completely abolish activity. This suggested that fluorinase activity is not dependent on any metal-cofactor. It is probable, however, that Mg^{2+} induces conformational changes to the native structure allowing for slightly enhanced activities. Similar findings have been reported for other proteins [134]. This observation can be verified using fluorescence spectroscopy and circular dichroism to observe secondary and tertiary structures of the enzyme in the presence and absence of Mg^{2+} [134,135]. It was also found that 26% of activity was retained after two months at 25 °C. It is not surprising that the *A. mzabensis* fluorinase displayed superior thermostability as the organism naturally thrives in higher temperatures [136]. Although, it would be interesting to assess the rate in which activity decreases over the two months. In addition, characterisation of the enzyme in terms of substrate scope, kinetic properties and structural studies are recommended. However, these preliminary characterisation studies revealed an enzyme with unique properties that can be potentially exploited for biocatalytic usage.

5.3 Conclusions

This study successfully produced the first comprehensive transcriptome for *D. cymosum*, significantly expanding genetic information available for this plant. Moreover, wounding studies and differential expression analysis provided insights into the molecular mechanisms that govern the wound response in *D. cymosum*. The proteome for various *D. cymosum* tissues were also sequenced and integrated with transcriptome data. Although the fluorinase from *D. cymosum* was not obtained using the aforementioned approaches, the data produced here will serve as a platform for future investigations into this plant.

A highly robust fluorinase was identified in the bacterium, *Actinopolyspora mzabensis*. A method for refolding of the enzyme was presented and, using this approach, we attained higher yields of the fluorinase. To the best of our knowledge, the successful refolding of a fluorinase enzyme has not been reported prior to this study and the method presented can be applied to other fluorinases. Preliminary characterisation studies provided the optimal conditions of the *A. mzabensis* fluorinase making it more amenable for practical usage. The identification of a novel fluorinase in this study significantly contributes to the biological tools available for biofluorination.

References

1. Ruff, O. and Krug, H. (1930) On a New Chlorofluoride ClF_3. *Journal of Inorganic and General Chemistry*, *190*, 270-276.
2. Fried, J. and Sabo, E.F. (1953) Synthesis of 17α-hydroxycorticosterone and its 9α-halo derivatives from 11-epi-17α-hydroxycorticosterone. *Journal of the American Chemical Society*, *75*, 2273-2274.
3. Fried, J. and Sabo, E.F. (1954) 9α-Fluoro derivatives of cortisone and hydrocortisone. *Journal of the American Chemical Society*, *76*, 1455-1456.
4. Wang, J., Sánchez-Roselló, M., Aceña, J.L., del Pozo, C., Sorochinsky, A.E., Fustero, S., Soloshonok, V.A. and Liu, H. (2013) Fluorine in pharmaceutical industry: fluorine-containing drugs introduced to the market in the last decade (2001–2011). *Chemical reviews*, *114*, 2432-2506.
5. Heidelberger, C., Chaudhuri, N. K., Danneberg, P., Mooren, D., Griesbach, L., Duschinsky, R., Schnitzer, R. J., Pleven, E. and J. Scheiner (1957) Fluorinated pyrimidines, a new class of tumour-inhibitory compounds. *Nature*, 179, 663.
6. Longley, D. B., Harkin, D. P., and Johnston, P. G. (2003) 5-fluorouracil: mechanisms of action and clinical strategies. *Nature Reviews Cancer*, *3*, 330.
7. Isanbor, C. and O'Hagan, D. (2006) Fluorine in medicinal chemistry: A review of anti-cancer agents. *Journal of Fluorine Chemistry*, *127*, 303-319.
8. Müller, K., Faeh, C., and Diederich, F. (2007) Fluorine in pharmaceuticals: looking beyond intuition. *Science*, *317*, 1881-1886.
9. Fujiwara, T. and O'Hagan, D. (2014) Successful fluorine-containing herbicide agrochemicals. *Journal of Fluorine Chemistry*, *167*, 16-29.
10. Purser, S., Moore, P. R., Swallow, S., and Gouverneur, V. (2008) Fluorine in medicinal chemistry. *Chemical Society Reviews*, *37*, 320-330.
11. Harper, D. B. and O'Hagan, D. (1994) The fluorinated natural products. *Natural Product Reports*, *11*, 123-133.
12. O'Hagan, D. (2008) Understanding organofluorine chemistry. An introduction to the C–F bond. *Chemical Society Reviews*, *37*, 308-319.
13. Smart, B. E. (2001) Fluorine substituent effects (on bioactivity). *Journal of Fluorine Chemistry*, *109*, 3-11.

14. Biffinger, J. C., Kim, H. W., and DiMagno, S. G. (2004) The polar hydrophobicity of fluorinated compounds. *ChemBioChem*, *5*, 622-627.
15. Danielson, P. B. (2002) The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. *Current Drug Metabolism*, *3*, 561-597.
16. Gribble, G. W. (2003) The diversity of naturally produced organohalogens. *Chemosphere*, *52*, 289-297.
17. Gribble, G. W. (2004) Natural organohalogens: a new frontier for medicinal agents? *Journal of Chemical Education*, *81*, 1441.
18. Gribble, G. W. (2012) The Alkaloids: Chemistry and Biology. *Academic Press*.
19. O'Hagan, D., and Deng, H. (2014) Enzymatic fluorination and biotechnological developments of the fluorinase. *Chemical Reviews*, *115*, 634-649.
20. Greenwood, N. N., and Earnshaw, A. (1997) *Chemistry of the Elements 2nd Edition*. Butterworth-Heinemann.
21. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., Garneau-Tsodikova, S., and Walsh, C. T. (2006) Nature's inventory of halogenation catalysts: oxidative strategies predominate. *Chemical Reviews*, *106*, 3364-3378.
22. Agarwal, V., Miles, Z. D., Winter, J. M., Eustáquio, A. S., El Gamal, A. A. and Moore, B. S. (2017) Enzymatic halogenation and dehalogenation reactions: pervasive and mechanistically diverse. *Chemical Reviews*, *117*, 5619-5674.
23. Shaw, P. D. and Hager, L. P. (1959) β-Ketoadipate chlorinase: a soluble enzyme system. *Journal of Biological Chemistry*, *234*, 2565-2569.
24. Shaw, P. D. and Hager, L. P. (1961) Chloroperoxidase: a component of the β-ketoadipate chlorinase system. *Journal of Biological Chemistry*, *236*, 1626-1630.
25. Hager, L. P., Morris, D. R., Brown, F. S. and Eberwein, H. (1966) Utilization of halogen anions. *Journal of Biological Chemistry*, *241*, 1769-1777.
26. Bernhardt, P., Okino, T., Winter, J. M., Miyanaga, A. and Moore, B. S. (2011) A stereoselective vanadium-dependent chloroperoxidase in bacterial antibiotic biosynthesis. *Journal of the American Chemical Society*, *133*, 4268-4270.
27. Hartung, J., Brücher, O., Hach, D., Schulz, H., Vilter, H. and Ruick, G. (2008) Bromoperoxidase activity and vanadium level of the brown alga *Ascophyllum nodosum*. *Phytochemistry*, *69*, 2826-2830.
28. Rorrer, G. L., Tucker, M. P., Cheney, D. P. and Maliakal, S. (2001) Bromoperoxidase activity in microplantlet suspension cultures of the macrophytic red alga *Ochtodes secundiramea*. *Biotechnology and Bioengineering*, *74*, 389-395.

29. Speicher, A., Heisel, R. and Kolz, J. (2003) First detection of a chloroperoxidase in bryophytes. *Phytochemistry*, *62*, 679-682.
30. Wagner, C., Molitor, I. M. and König, G. M. (2008) Critical view on the monochlorodimedone assay utilized to detect haloperoxidase activity. *Phytochemistry*, *69*, 323-332.
31. Messerschmidt A. and Wever R. (1996) X-ray structure of a vanadium-containing enzyme: chloroperoxidase from the fungus *Curvularia inaequalis*. *Proceedings of the National Academy of Science, 93*, 392–396.
32. Jacks, T. J. (2005) Evaluation of kojic acid for determining heme and nonheme haloperoxidase activities spectrofluorometrically. *Analytical Letters*, *38*, 921-927.
33. Sundaramoorthy, M., Terner, J. and Poulos, T. L. (1995) The crystal structure of chloroperoxidase: a heme peroxidase–cytochrome P450 functional hybrid. *Structure*, *3*, 1367-1378.
34. Bitto, E., Huang, Y., Bingman, C. A., Singh, S., Thorson, J. S. and Phillips, G. N. (2008) The structure of flavin-dependent tryptophan 7-halogenase RebH. *Proteins: Structure Function and Bioinformatics, 70*, 289–293
35. Podzelinska, K., Latimer, R., Bhattacharya, A., Vining, L. C., Zechel, D. L. and Jia, Z. (2010) Chloramphenicol biosynthesis: the structure of CmlS, a flavin-dependent halogenase showing a covalent flavin–aspartate bond. *Journal of Molecular Biology*, *397*, 316-331.
36. Galonić, D. P., Vaillancourt, F. H. and Walsh, C. T. (2006) Halogenation of unactivated carbon centers in natural product biosynthesis: trichlorination of leucine during barbamide biosynthesis. *Journal of the American Chemical Society*, *128*, 3900-3901.
37. Vaillancourt, F. H., Yin, J. and Walsh, C. T. (2005) SyrB2 in syringomycin E biosynthesis is a nonheme Fe(II) alpha-ketoglutarate- and O2-dependent halogenase. *Proceedings of the National Academy of Sciences USA*, 102, 10111–10116.
38. Schmidberger, J. W., James, A. B., Edwards, R., Naismith, J. H. and O'Hagan D. (2010) Halomethane biosynthesis: structure of a SAM-dependent halide methyltransferase from *Arabidopsis thaliana*. *Angewandte Chemie*. *49*, 3646–3648.
39. Eustáquio, A. S., Pojer, F., Noel, J. P. and Moore, B. S. (2008) Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nature Chemical Biology*, *4*, 69.
40. O'Hagan, D., Schaffrath, C., Cobb, S. L., Hamilton, J. T. and Murphy, C. D. (2002) Biochemistry: biosynthesis of an organofluorine molecule. *Nature*, *416*, 279.

41. Saxena, D., Aouad, S., Attieh, J. and Saini, H. S. (1998) Biochemical characterization of chloromethane emission from the wood-rotting fungus *Phellinus pomaceus*. *Applied and Environmental Microbiology, 64*, 2831-2835.
42. Ohsawa, N., Tsujita, M., Morikawa, S. and Itoh, N. (2001) Purification and characterization of a monohalomethane-producing enzyme *S*-adenosyl-L-methionine: halide ion methyltransferase from a marine microalga, *Pavlova pinguis*. *Bioscience, Biotechnology, and Biochemistry*, *65*, 2397-2404.
43. Attieh, J., Sparace, S. A. and Saini, H. S. (2000) Purification and properties of multiple isoforms of a novel thiol methyltransferase involved in the production of volatile sulfur compounds from *Brassica oleracea*. *Archives of Biochemistry and Biophysics*, *380*, 257-266.
44. Itoh, N., Toda, H., Matsuda, M., Negishi, T., Taniguchi, T. and Ohsawa, N. (2009) Involvement of *S*-adenosylmethionine-dependent halide/thiol methyltransferase (HTMT) in methyl halide emissions from agricultural plants: isolation and characterization of an HTMT-coding gene from *Raphanus sativus* (daikon radish). *BMC Plant Biology*, *9*, 116.
45. Dong, C., Huang, F., Deng, H., Schaffrath, C., Spencer, J. B., O'Hagan, D. and Naismith, J. H. (2004) Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature*, *427*, 561.
46. De Sousa Correia, R. I. and Van Rensburg, L. (2000). Geographical distribution and local occurrence of *Dichapetalum cymosum* (Hook) Engl. (gifblaar) in Namibia and surrounding areas. *Agricola*, *11*, 19-28.
47. Kellerman, T. S., Naudé, T. W. and Fourie, N. (1996) The distribution, diagnoses and estimated economic impact of plant poisonings and mycotoxicoses in South Africa. *Onderstepoort Journal of Veterinary Research*. 63:65-69.
48. Nel, E., Robbertse, P. J. and Grobbelaar, N. (1982) A morphological study of *Dichapetalum cymosum* (Hook.) Engl. Part 1. The underground stem and root systems. *South African Journal of Botany*, *1*, 14-17.
49. Marais, J. S. C. (1943) The isolation of the toxic principle "potassium cymonate" from "Gifblaar" *Dichapetalum cymosum* (Hook) Engl. *Onderstepoort Journal of Veterinary Science and Animal Industry*, *18*, 203-206.
50. Marais, J. S. C. (1944) Monofluoroacetic acid, the toxic principle of "gifblaar", *Dichapetalum cymosum* (Hook) Engl. *Onderstepoort Journal of Veterinary Science and Animal Industry*, *20*, 67-73.

51. Meyer, J. M. and Grobbelaar, N. (1990) The determination, uptake and transport of fluoroacetate in *Dichapetalum cymosum*. *Journal of Plant Physiology*, *135*, 546-549.
52. Tannock, J. (1975) Seasonal variation in toxicity of *Dichapetalum cymosum* (Hook) Engl. in the Nyamandhlovu area. *Rhodesia, Zambia and Malawi Journal of Agricultural Research*.
53. Goncharov, N. V., Jenkins, R. O. and Radilov, A. S. (2006) Toxicology of fluoroacetate: a review, with possible directions for therapy research. *Journal of Applied Toxicology*, *26*, 148-161.
54. Peters, R. A., Wakelin, R. W., Buffa, P. and Thomas, L. C. (1953) Biochemistry of fluoroacetate poisoning. The isolation and some properties of the fluorotricarboxylic acid inhibitor of citrate metabolism. *Proceedings of the Royal Society of London. Series B-Biological Sciences*, *140*, 497-506.
55. Morrison, J. F. and Peters, R. A. (1954) Biochemistry of fluoroacetate poisoning: the effect of fluorocitrate on purified aconitase. *Biochemical Journal*, *58*, 473.
56. Treble, D. H., Lamport, D. T. A. and Peters, R. A. (1962) The inhibition of plant aconitate hydratase (aconitase) by fluorocitrate. *Biochemical Journal*, *85*, 113.
57. Marcus, A. and Elliott, W. B. (1956) Enzymatic reactions of fluoroacetate and fluoroacetyl coenzyme A. *Journal of Biological Chemistry*, *218*, 823-830.
58. Lauble, H., Kennedy, M. C., Emptage, M. H., Beinert, H. and Stout, C. D. (1996) The reaction of fluorocitrate with aconitase and the crystal structure of the enzyme-inhibitor complex. *Proceedings of the National Academy of Sciences*, *93*, 13699-13703.
59. Buffa, P. and Peters, R. A. (1949) The *in vivo* formation of citrate induced by fluoroacetate and its significance. *The Journal of Physiology*, *110*, 488-500.
60. Kun, E., Kirsten, E. and Sharma, M. L. (1977) Enzymatic formation of glutathione-citryl thioester by a mitochondrial system and its inhibition by (-) erythrofluorocitrate. *Proceedings of the National Academy of Sciences*, *74*, 4942-4946.
61. Chenoweth, M. B. and Gilman, A. (1946) Studies on the pharmacology of fluoroacetate: I. Species response to fluoroacetate. *Journal of Pharmacology and Experimental Therapeutics*, *87*, 90-103.
62. Nwude, N., Parsons, L. E. and Adaudi, A. O. (1977) Acute toxicity of the leaves and extracts of *Dichapetalum barteri* (Engl.) in mice, rabbits and goats. *Toxicology*, *7*, 23-29.

63. Harrisson, J. W., Ambrus, J. L., Ambrus, C. M., Rees, E. W., Peters, R. H., Reese, L. C. and Baker, T. (1952) Acute poisoning with sodium fluoroacetate (compound 1080). *Journal of the American Medical Association*, *149*, 1520-1522.
64. Chung, H. M. (1984) Acute renal failure caused by acute monofluoroacetate poisoning. *Veterinary and Human Toxicology*, *26*, 29-32.
65. O'Hagan, D., Perry, R., Lock, J. M., Meyer, J. M., Dasaradhi, L., Hamilton, J. T. and Harper, D. B. (1993) High levels of monofluoroacetate in *Dichapetalum braunii*. *Phytochemistry*, *33*, 1043-1045.
66. Vickery, B. and Vickery, M. L. (1972) Fluoride metabolism in *Dichapetalum toxicarium*. *Phytochemistry*, *11*, 1905-1909.
67. Vickery, B., Vickery, M. L. and Ashu, J. T. (1973) Analysis of plants for fluoroacetic acids. *Phytochemistry*, *12*, 145-147.
68. Vickery, B. and Vickery, M. L. (1975) The synthesis and defluorination of monofluoroacetate in some *Dichapetalum* species. *Phytochemistry*, *14*, 423-427.
69. Esters, V., Karangwa, C., Tits, M., Francotte, P., Pirotte, B., Servais, A.C., Fillet, M., Crommen, J., Robbrecht, E., Minet, A. and Grisar, T. (2013) Unusual amino acids and monofluoroacetate from *Dichapetalum michelsonii* (Umutambasha), a toxic plant from Rwanda. *Planta Medica*, *79*, 334-337.
70. Hall, R. J. (1972) The distribution of organic fluorine in some toxic tropical plants. *New Phytologist*, *71*, 855-871.
71. Hamilton, J. T. and Harper, D. B. (1997) Fluoro fatty acids in seed oil of *Dichapetalum toxicarium*. *Phytochemistry*, *44*, 1129-1132.
72. Christie, W. W., Hamilton, J. T. G. and Harper, D. B. (1998) Mass spectrometry of fluorinated fatty acids in the seed oil of *Dichapetalum toxicarium*. *Chemistry and Physics of Lipids*, *97*, 41-47.
73. Kellerman, T. S., Coetzer, J. A. W., Naudé, T. W. and Botha, C. J. (2005) *Plant Poisonings and Mycotoxicoses of Livestock in Southern Africa.* (No. Ed. 2). Oxford University Press, Southern Africa.
74. Leong, L. E. X., Khan, S., Davis, C. K., Denman, S. E. and McSweeney, C. S. (2017) Fluoroacetate in plants-a review of its distribution, toxicity to livestock and microbial detoxification. *Journal of Animal Science and Biotechnology*, *8*, 55.
75. Oelrichs, P. B. and McEwan, T. (1961) Isolation of the toxic principle in *Acacia georginae*. *Nature*, *190*, 808.

76. Twigg, L. E., King, D. R., Bowen, L. H., Wright, G. R. and Eason, C. T. (1996) Fluoroacetate content of some species of the toxic Australian plant genus, *Gastrolobium,* and its environmental persistence. *Natural Toxins*, *4*, 122-127.
77. Twigg, L. E., Wright, G. R. and Potts, M. D. (1999) Fluoroacetate content of *Gastrolobium brevipes* in central Australia. *Australian Journal of Botany*, *47*, 877-880.
78. Baron, M. L., Bothroyd, C. M., Rogers, G. I., Staffa, A. and Rae, I. D. (1987) Detection and measurement of fluoroacetate in plant extracts by ^{19}F-NMR. *Phytochemistry*, *26*, 2293-2295.
79. Peacock, D. E., Williams, B. D. and Christensen, P. E. (2007) Total fluorine analysis of seed of Australian *Gastrolobium spp*. showing temporal, spatial and morphological variation. *Journal of Fluorine Chemistry*, *128*, 631-635.
80. Lee, S. T., Cook, D., Pfister, J. A., Allen, J. G., Colegate, S. M., Riet-Correa, F. and Taylor, C. M. (2014) Monofluoroacetate-containing plants that are potentially toxic to livestock. *Journal of Agricultural and Food Chemistry*, *62*, 7345-7354.
81. Tokarnia, C. H., Döbereiner, J. and Peixoto, P. V. (2002) Poisonous plants affecting livestock in Brazil. *Toxicon*, *40*, 1635-1660.
82. Krebs, H. C., Kemmerling, W. and Habermehl, G. (1994) Qualitative and quantitative determination of fluoroacetic acid in *Arrabidea bilabiata* and *Palicourea marcgravii* by ^{19}F-NMR spectroscopy. *Toxicon, 32*, 909-913.
83. Lee, S. T., Cook, D., Riet-Correa, F., Pfister, J. A., Anderson, W. R., Lima, F. G. and Gardner, D. R. (2012) Detection of monofluoroacetate in *Palicourea* and *Amorimia* species. *Toxicon*, *60*, 791-796.
84. Cook, D., Lee, S. T., Taylor, C. M., Bassüner, B., Riet-Correa, F., Pfister, J. A. and Gardner, D. R. (2014) Detection of toxic monofluoroacetate in *Palicourea* species. *Toxicon*, *80*, 9-16.
85. Hall, R. J. and Cain, R. B. (1972) Organic fluorine in tropical soils. *New Phytologist*, *71*, 839-853.
86. Meyer, J. J., Grobbelaar, N. and Steyn, P. L. (1990) Fluoroacetate-metabolizing pseudomonad isolated from *Dichapetalum cymosum*. *Applied and Environmental Microbiology*, *56*, 2152-2155.
87. Grobbelaar, N. and Meyer, J. M. (1990) Fluoroacetate production by *Dichapetalum cymosum*. *Journal of Plant Physiology*, *135*, 550-553.

88. Preuss, P. W., Colavito, L. T. and Weinstein, L. H. (1970) The synthesis of monofluoroacetic acid by a tissue culture of *Acacia georginae*. *Cellular and Molecular Life Sciences*, *26*, 1059-1060.
89. Meyer, J. M. and O'Hagan, D. (1992) Conversion of fluoropyruvate to fluoroacetate by *Dichapetalum cymosum*. *Phytochemistry*, *31*, 499-501.
90. Meyer, J. M. and O'Hagan, D. (1992) Conversion of 3-fluoropyruvate to fluoroacetate by cell-free extracts of *Dichapetalum cymosum*. *Phytochemistry*, *31*, 2699-2701.
91. Sanada, M., Miyano, T., Iwadare, S., Williamson, J. M., Arison, B. H., Smith, J. L., Douglas, A. W., Liesch, J. M. and Inamine, E., (1986). Biosynthesis of fluorothreonine and fluoroacetic. *The Journal of Antibiotics*, *39*, 259-265.
92. Reid, K. A., Bowden, R. D., Dasaradhi, L., Amin, M. R. and Harper, D. B. (1995) Biosynthesis of fluorinated secondary metabolites by *Streptomyces cattleya*. *Microbiology*, *141*, 1385-1393.
93. Schaffrath, C., Murphy, C. D., Hamilton, J. T. and O'Hagan, D. (2001) Biosynthesis of fluoroacetate and 4-fluorothreonine in *Streptomyces cattleya*. Incorporation of oxygen-18 from [2-^{2} H, 2-^{18} O]-glycerol and the role of serine metabolites in fluoroacetaldehyde biosynthesis. *Journal of the Chemical Society,* 23, 3100-3105.
94. Moss, S. J., Murphy, C. D., Hamilton, J. T., McRoberts, W. C., O'Hagan, D., Schaffrath, C. and Harper, D. B. (2000) Fluoroacetaldehyde: a precursor of both fluoroacetate and 4-fluorothreonine in *Streptomyces cattleya*. *Chemical Communications*, 22, 2281-2282.
95. Murphy, C. D., Moss, S. J. and O'Hagan, D. (2001) Isolation of an Aldehyde Dehydrogenase Involved in the Oxidation of Fluoroacetaldehyde to Fluoroacetate in *Streptomyces cattleya*. *Applied and Environmental Microbiology*, *67*, 4919-4921.
96. Murphy, C. D., O'Hagan, D. and Schaffrath, C. (2001) Identification of a PLP-dependent threonine transaldolase: a novel enzyme involved in 4-fluorothreonine biosynthesis in *Streptomyces cattleya*. *Angewandte Chemie*, *40*, 4479-4481.
97. Deng, H., O'Hagan, D. and Schaffrath, C. (2004) Fluorometabolite biosynthesis and the fluorinase from *Streptomyces cattleya*. *Natural Product Reports*, *21*, 773-784.
98. Schaffrath, C., Cobb, S. L. and O'Hagan, D. (2002) Cell-free biosynthesis of fluoroacetate and 4-fluorothreonine in *Streptomyces cattleya*. *Angewandte Chemie*, *41*, 3913-3915.

99. Schaffrath, C., Deng, H. and O'Hagan, D. (2003) Isolation and characterisation of 5′-fluorodeoxyadenosine synthase, a fluorination enzyme from *Streptomyces cattleya*. *FEBS Letters*, *547*, 111-114.
100. Cobb, S. L., Deng, H., Hamilton, J. T., McGlinchey, R. P. and O'Hagan, D. (2004) Identification of 5-fluoro-5-deoxy-D-ribose-1-phosphate as an intermediate in fluorometabolite biosynthesis in *Streptomyces cattleya*. *Chemical Communications*, 5, 592-593.
101. Onega, M., McGlinchey, R. P., Deng, H., Hamilton, J. T. and O'Hagan, D. (2007) The identification of (3*R*, 4*S*)-5-fluoro-5-deoxy-D-ribulose-1-phosphate as an intermediate in fluorometabolite biosynthesis in *Streptomyces cattleya*. *Bioorganic Chemistry*, *35*(5), 375-385.
102. Deng, H., Cross, S. M., McGlinchey, R. P., Hamilton, J. T. and O'Hagan, D. (2008) *In vitro* reconstituted biotransformation of 4-fluorothreonine from fluoride ion: application of the fluorinase. *Chemistry & Biology*, *15*, 1268-1276.
103. Zhao, C., Li, P., Deng, Z., Ou, H. Y., McGlinchey, R. P. and O'Hagan, D. (2012) Insights into fluorometabolite biosynthesis in *Streptomyces cattleya* DSM46488 through genome sequence and knockout mutants. *Bioorganic Chemistry*, *44*, 1-7.
104. Zhu, X., Robinson, D. A., McEwan, A. R., O'Hagan, D. and Naismith, J. H. (2007) Mechanism of Enzymatic Fluorination in *Streptomyces cattleya*. *Journal of the American Chemical Society*, *129*, 14597-14604.
105. Barbe, V., Bouzon, M., Mangenot, S., Badet, B., Poulain, J., Segurens, B., Vallenet, D., Marlière, P. and Weissenbach, J. (2011) Complete genome sequence of *Streptomyces cattleya* NRRL 8057, a producer of antibiotics and fluorometabolites.
106. Huang, F., Haydock, S.F., Spiteller, D., Mironenko, T., Li, T.L., O'Hagan, D., Leadlay, P.F. and Spencer, J.B. (2006) The gene cluster for fluorometabolite biosynthesis in *Streptomyces cattleya*: A thioesterase confers resistance to fluoroacetyl-coenzyme A. *Chemistry & Biology*, *13*, 475-484.
107. Weeks, A. M., Coyle, S. M., Jinek, M., Doudna, J. A. and Chang, M. C. (2010) Structural and biochemical studies of a fluoroacetyl-CoA-specific thioesterase reveal a molecular basis for fluorine selectivity. *Biochemistry*, *49*, 9269-9279.
108. Deng, H., Ma, L., Bandaranayaka, N., Qin, Z., Mann, G., Kyeremeh, K., Yu, Y., Shepherd, T., Naismith, J.H. and O'Hagan, D. (2014) Identification of fluorinases from *Streptomyces* sp MA37, *Norcardia brasiliensis*, and *Actinoplanes* sp N902-109 by Genome Mining. *ChemBioChem*, *15*, 364-368.

109. Wang, Y., Deng, Z. and Qu, X. (2014) Characterization of a SAM-dependent fluorinase from a latent biosynthetic pathway for fluoroacetate and 4-fluorothreonine formation in *Nocardia brasiliensis*. *F1000Research*, *3*.
110. Ma, L., Bartholome, A., Tong, M.H., Qin, Z., Yu, Y., Shepherd, T., Kyeremeh, K., Deng, H. and O'Hagan, D. (2015). Identification of a fluorometabolite from *Streptomyces* sp. MA37:(2*R*, 3*S*, 4*S*)-5-fluoro-2,3,4-trihydroxypentanoic acid. *Chemical Science*, *6*, 1414-1419.
111. Zhao, X. and Yang, T. (2011) Draft genome sequence of the marine sediment-derived actinomycete *Streptomyces xinghaiensis* NRRL B24674T.
112. Ma, L., Li, Y., Meng, L., Deng, H., Li, Y., Zhang, Q. and Diao, A. (2016) Biological fluorination from the sea: discovery of a SAM-dependent nucleophilic fluorinating enzyme from the marine-derived bacterium *Streptomyces xinghaiensis* NRRL B24674. *RSC Advances*, *6*, 27047-27051.
113. HimáTong, M. (2014) Fluoroacetate biosynthesis from the marine-derived bacterium *Streptomyces xinghaiensis* NRRL B-24674. *Organic & Biomolecular Chemistry*, *12*, 4828-4831.
114. Thomas, S.O., Singleton, V.L., Lowery, J.A., Sharpe, R.W., Pruess, L.M., Porter, J.N., Mowat, J.H. and Bohonos, N. (1957) Nucleocidin, a new antibiotic with activity against Trypanosomes. *Antibiotics Annual*, 716-721.
115. Morton, G. O., Lancaster, J. E., Van Lear, G. E., Fulmor, W. and Meyer, W. E. (1969) Structure of nucleocidin. III. Revised structure. *Journal of the American Chemical Society*, *91*, 1535-1537.
116. Kalan, L., Gessner, A., Thaker, M.N., Waglechner, N., Zhu, X., Szawiola, A., Bechthold, A., Wright, G.D. and Zechel, D.L. (2013) A cryptic polyene biosynthetic gene cluster in *Streptomyces calvus* is expressed upon complementation with a functional *bldA* gene. *Chemistry & Biology*, *20*, 1214-1224.
117. Zhu, X.M., Hackl, S., Thaker, M.N., Kalan, L., Weber, C., Urgast, D.S., Krupp, E.M., Brewer, A., Vanner, S., Szawiola, A. and Yim, G. (2015) Biosynthesis of the Fluorinated Natural Product Nucleocidin in *Streptomyces calvus* Is Dependent on the *bldA*-Specified Leu-tRNA^{UUA} Molecule. *ChemBioChem*, *16*, 2498-2506.
118. Bartholomé, A., Janso, J. E. and O'Hagan, D. (2017) Fluorometabolite biosynthesis: isotopically labelled glycerol incorporations into the antibiotic nucleocidin in *Streptomyces calvus*. *Organic & Biomolecular Chemistry*, *15*, 61-64.

119. Liang, T., Neumann, C. N. and Ritter, T. (2013) Introduction of Fluorine and Fluorine-Containing Functional Groups. *Angewandte Chemie*, *52*, 8214-8264.
120. Truppo, M. D. (2017) Biocatalysis in the pharmaceutical industry: the need for speed. *ACS Medicinal Chemistry Letters*, *8*, 476-480.
121. Martarello, L., Schaffrath, C., Deng, H., Gee, A. D., Lockhart, A. and O'Hagan, D. (2003) The first enzymatic method for C–^{18}F bond formation: the synthesis of 5′-[^{18}F]-fluoro-5′-deoxyadenosine for imaging with PET. *Journal of Labelled Compounds and Radiopharmaceuticals: The Official Journal of the International Isotope Society*, *46*, 1181-1189.
122. Li, X. G., Domarkas, J. and O'Hagan, D. (2010) Fluorinase mediated chemoenzymatic synthesis of [^{18}F]-fluoroacetate. *Chemical Communications*, *46*, 7819-7821.
123. Winkler, M., Domarkas, J., Schweiger, L. F. and O'Hagan, D. (2008) Fluorinase-coupled base swaps: synthesis of [^{18}F]-5′-deoxy-5′-fluorouridines. *Angewandte Chemie*, *47*, 10141-10143.
124. Dall'Angelo, S., Bandaranayaka, N., Windhorst, A.D., Vugts, D.J., van der Born, D., Onega, M., Schweiger, L.F., Zanda, M. and O'Hagan, D. (2013) Tumour imaging by Positron Emission Tomography using fluorinase generated 5-[^{18}F] fluoro-5-deoxyribose as a novel tracer. *Nuclear Medicine and Biology*, *40*, 464-470.
125. Eustáquio, A. S., O'Hagan, D. and Moore, B. S. (2010) Engineering fluorometabolite production: fluorinase expression in *Salinispora tropica* yields fluorosalinosporamide. *Journal of Natural Products*, *73*, 378-382.
126. Hong, H., Spiteller, D. and Spencer, J. B. (2008) Incorporation of fluoroacetate into an aromatic polyketide and its influence on the mode of cyclization. *Angewandte Chemie*, *47*, 6028-6032.
127. Sturmberger, L., Wallace, P. W., Glieder, A. and Birner-Gruenberger, R. (2016) Synergism of proteomics and mRNA sequencing for enzyme discovery. *Journal of Biotechnology*, *235*, 132-138.
128. Gregory, T.R., Nicol, J.A., Tamm, H., Kullman, B., Kullman, K., Leitch, I.J., Murray, B.G., Kapraun, D.F., Greilhuber, J. and Bennett, M.D. (2006) Eukaryotic genome size databases. *Nucleic Acids Research*, *35*, D332-D338.
129. Schatz, M. C., Witkowski, J. and McCombie, W. R. (2012) Current challenges in *de novo* plant genome sequencing and assembly. *Genome Biology*, *13*, 243.

130. Yang, D., Du, X., Yang, Z., Liang, Z., Guo, Z. and Liu, Y. (2014) Transcriptomics, proteomics, and metabolomics to reveal mechanisms underlying plant secondary metabolism. *Engineering in Life Sciences*, *14*, 456-466.
131. Tang, Q., Ma, X., Mo, C., Wilson, I.W., Song, C., Zhao, H., Yang, Y., Fu, W. and Qiu, D. (2011) An efficient approach to finding *Siraitia grosvenorii* triterpene biosynthetic genes by RNA-seq and digital gene expression analysis. *BMC Genomics*, *12*, 343.
132. Shi, C.Y., Yang, H., Wei, C.L., Yu, O., Zhang, Z.Z., Jiang, C.J., Sun, J., Li, Y.Y., Chen, Q., Xia, T. and Wan, X.C. (2011) Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC Genomics*, *12*, 131.
133. Aslam, B., Basit, M., Nisar, M. A., Khurshid, M. and Rasool, M. H. (2017) Proteomics: technologies and their applications. *Journal of Chromatographic Science*, *55*, 182-196.
134. Kumar, S., Jain, K. K., Singh, A., Panda, A. K. and Kuhad, R. C. (2015) Characterization of recombinant pectate lyase refolded from inclusion bodies generated in *E. coli* BL21 (DE3). *Protein Expression and Purification*, *110*, 43-51.
135. Keshav, V., Achilonu, I., Dirr, H. W. and Kondiah, K. (2019) Recombinant expression and purification of a functional bacterial metallo-chaperone PbrD-fusion construct as a potential biosorbent for Pb (II). *Protein Expression and Purification*, *158*, 27-35.
136. Meklat, A., Bouras, N., Zitouni, A., Mathieu, F., Lebrihi, A., Schumann, P., Spröer, C., Klenk, H.P. and Sabaou, N. (2013). *Actinopolyspora mzabensis* sp. nov., a halophilic actinomycete isolated from an Algerian Saharan soil. *International Journal of Systematic and Evolutionary Microbiology*, *63*,3787-3792.